

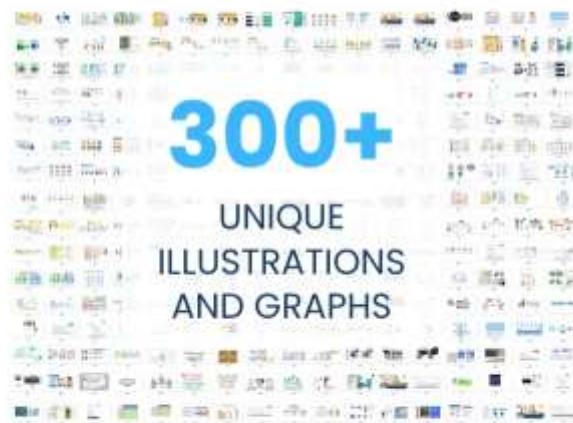
2. VYDANIE  
AKTUALIZOVANÉ A  
ROZŠÍRENÉ

# DATA DRIVEN CONSTRUCTION

NAVIGÁCIA V DÁTOVOM VEKU V  
STAVEBNÍCTVE

S PRÍPADMI  
POUŽITIA AI A LLM

Artem Boiko



# **DATA-DRIVEN CONSTRUCTION**

**Navigácia v dobe dátovej  
v stavebníctve**

Druhé, prepracované a doplnené vydanie

**ARTEM BOIKO**

**“**

"Bojko je James Carville v oblasti IT - v jeho často citovanom výroku "It's the economy, stupid" (Je to ekonomika, hlupáci) stačí vymeniť len jedno slovo za túto slávnu knihu. "Sú to údaje, hlupáci." (nie softvér) A na to, aby sme sa v dátovom vesmíre zorientovali, dodnes platí výrok starých Rimanov pochádzajúci z gréčtiny: "Navigare necesse est". Autor naviguje svojich čitateľov po všetkých hlinách a plytčinách dátového oceánu istou rukou a neochvějným kompasom, nehovoriac o komplexnom historickom prístupe a v neposlednom rade o veľmi originálnej grafike a dobrovzbudzujúcom zmysle pre humor, ktorý nie je viditeľný len na druhý pohľad. Medzinárodné ohlasy na Bojkovu knihu sa pohybujú od euforického súhlasu až po skôr žlčovitú skepsu, čo druhému nemeckému vydaniu knihy prospelo. Bojko je originálny a nedogmatický mysliteľ dát. Predkladá čitateľovi vzrušujúce postrehy a vždy odvážne, ba až provokatívne tézy, ktoré inšpirujú k ďalšiemu uvažovaniu. Výborný liek na nemeckú chorobu latentného konsenzualizmu. Mimochodom, uvedené latinské príslovie má aj doplnenie: "vivere non est necesse". Na Bojkov prístup k svetu dát sa nevzťahuje - dáta žijú a ich život je potrebný, nieto ešte rozhodujúci."

- **Dr. Burkhard Talebitari**, nezávislý redaktor - aj pre časopis: BIM, ktorý od roku 2013 každoročne vydáva spoločnosť Ernst & Sohn.

"Kniha Artema Boika je milníkom pre demokratizáciu digitalizácie v stavebníctve - a skutočnou zmenou pre malé a stredné podniky (MSP). Je obzvlášť prelomová: pomocou moderných open-source nástrojov s nízkym kódom a bez kódu môžu podniky už teraz efektívne integrovať údaje do svojich podnikových procesov a výhodne ich analyzovať - bez akýchkoľvek hlbších znalostí programovania. Tým sa stáva nákladné používanie tažkopádnych komerčných softvérových balíkov zbytočným. Táto kniha je výzvou k akcii! Je cennou príručkou pre každého, kto chce nielen pochopiť digitálnu transformáciu v stavebníctve, ale chce ju aj aktívne formovať - pragmaticky, efektívne a perspektívne. Teraz je čas spolupracovať na zdieľaní týchto poznatkov a trvalo zvyšovať produktivitu stavebníctva."

- **Dr. Michael Max Buehler**, profesor stavebného manažmentu na HTWG Konstanz, spolumajiteľ spoločnosti GemeinWerk Ventures a nezávislý riaditeľ spoločnosti DevvStream.

"Kniha DataDrivenConstruction je jedným z prvých krovok za hranice bežného sveta stavbárov s ich zložitými systémami projektovania a riadenia, keď, zdá sa, zložitosť a presýtenosť dát nedáva šancu ani na radikálne zjednodušenie a zvýšenie transparentnosti práce so stavebnými dátami. Artem vo svojej knihe jednoduchým jazykom ukazuje, aké možnosti pred nami otvárajú moderne technológie práce s dátami, a doslova uvádza konkrétné kroky, ktoré môžete okamžite aplikovať vo svojej práci. Vyzývam všetkých, ktorí chcú pochopiť, kam sa budú uberať automatizačné systémy v stavebníctve, aby si túto knihu pozorne preštudovali a uvedomili si, že dátová revolúcia v stavebníctve nám už klope na dvore. Teraz je zaujímavá len pre geekov, ale o niekoľko rokov, podobne ako BIM, budú takéto prístupy a softvér všadeprítomné!"

- **Ihor Rogachev**, vedúci kompetenčného centra IMT, BIM a digitálnej transformácie v RGD a zakladateľ InfraBIM.Pro.

"Vrelo odporúčam knihu DataDrivenConstruction, ktorá sa zaobera, ako hovorí názov, prístupom k riadeniu informácií na základe údajov pre spoločnosť AECO. V súčasnosti ju používam na pomoc pri iniciovaní viacerých diskusií s rôznymi skupinami. Zistil som, že je to veľmi prístupná referencia. Okrem dôkladného prehľadu historických súvislostí nástrojov v AECO, údajov a predstavenia niekoľkých klúčových technológií kniha obsahuje množstvo veľmi užitočných schém, ktoré načrtávajú rozsah zdrojov údajov a artefaktov pre koncových používateľov so vzorovými pracovnými postupmi. Zdá sa mi, že práve tieto typy diagramov potrebujeme viac pri tvorbe a monitorovaní informačných stratégii a prispievajú k BEP - definujú celkový dátový model podniku, na ktorý možno prekryť hranice pre PIM a AIM."

- **Paul Ransley**, hlavný konzultant spoločnosti Acmena a inžinier systémovej integrácie v spoločnosti Transport for London.

"Ak sú "dáta novou ropou", musíme sa ich naučiť definovať, nájsť, vyťažiť, zušľachtiť, aby boli hodnotné. Knihu DataDrivenConstruction považujem za veľmi poučnú a prenikavú. Kniha poskytuje užitočné historické pozadie a vysvetluje prácu s údajmi zrozumiteľným jazykom. Pre tých, ktorí sa zaujímajú o digitálnu transformáciu, poskytuje dobré pochopenie údajov - ako fungujú, ako sú štruktúrované a ako ich možno využiť."

- **Ralph Montague**, riaditeľ spoločnosti ArcDox, riaditeľ sumitu koordinátorov BIM a predseda národného zrakadlového výboru BIM v írskom národnom normalizačnom orgáne.

"Ako bolo v knihe zdôraznené, informácie sú pre stavebný sektor klúčovým prínosom a ich dostupnosť v prístupných formátoch výrazne uľahčuje presné rozhodovanie a urýchľuje časový harmonogram projektu. Kniha ponúka neutrálny a efektívny prístup k prístupu k tomuto zdroju a jeho využitiu pri rozhodovaní. Metodika prezentovaná v knihe využíva moderný prístup, ktorý kombinuje programovanie založené na umelej inteligencii s dostupnými nástrojmi s otvoreným zdrojovým kódom. Využitím sily umelej inteligencie a využitím softvéru s otvoreným zdrojovým kódom je cieľom metodiky zvýšiť automatizáciu, optimalizovať procesy a podporiť dostupnosť a spoluprácu v tejto oblasti. Jazyk knihy je zrozumiteľný a ľahko pochopiteľný."

- **Dr. Salih Ofluoğlu**, dekan Fakulty výtvarných umení a architektúry na Univerzite Antalya Bilim a organizátor Eurázijského fóra BIM.

"Všetko, čo môžem povedať je: WOW! Spôsob, akým ste zakomponovali história, LLM, grafiku a celkovú jednoduchosť pochopenia vašich bodov, je naozaj pozoruhodný. Priebeh knihy je úžasný. V tejto knihe je toľko brilantrých aspektov; skutočne mení pravidlá hry. Je to skvelý zdroj informácií a chválim vás za úsilie a väšeň, ktoré ste do nej vložili. Gratulujem k vytvoreniu takého pozoruhodného diela. Mohol by som pokračovať, ale stačí povedať, že som neuveriteľne ohromený!"

- **Natasha Prinsloo**, vedúca digitálnej praxe v spoločnosti energylab\_

"Pre každého v stavebníctve, od nováčikov až po skúsených profesionálov, táto kniha mení pravidlá hry! Nie je to typické zaprášené čítanie - je plná postrehov, stratégii a štipky humoru, ktoré vás zaujmú. Pokrýva vývoj využívania údajov v stavebníctve - od starodávnych metód zaznamenávania údajov až po najmodernejšie digitálne technológie. Je to ako preletieť strojom času vývojom údajov v stavebníctve. Či už ste architekt, inžinier, projektový manažér alebo dátový analytik, táto komplexná príručka revolučným spôsobom zmení váš prístup k projektom. Pripravte sa na optimalizáciu procesov, zlepšenie rozhodovania a riadenie projektov ako nikdy predtým!"

- **Pierpaolo Vergati**, prednášajúci na univerzite Sapienza v Ríme a vedúci manažér stavebných projektov v spoločnosti Fintecna.

"Knihu som prečítať jedným dychom za menej ako 6 hodín. Kvalita spracovania knihy je výborná, hustý lesklý papier, farebné prevedenie, príjemné písmo. Veľké množstvo praktických príkladov na prácu s LLM špecifických pre stavebníctvo vám ušetrí mesiace, ak nie roky samoštúdia. Príklady práce sú veľmi rôznorodé, od jednoduchých až po zložité, bez toho, aby ste museli kupovať zložitý a drahý softvér. Kniha umožní majiteľom akéhokoľvek podniku v stavebníctve pozrieť sa na svoju obchodnú stratégiu, digitalizáciu a perspektívnu rozvoja novým pohľadom. A menším spoločnostiam zvýšiť efektivitu pomocou cenovo dostupných a bezplatných nástrojov."

- **Mikhail Kosarev**, lektor a konzultант pre digitálnu transformáciu v stavebníctve v spoločnosti TIM-ASG.

Kniha "DATA DRIVEN CONSTRUCTION" mení pravidlá hry pre každého, kto sa zaujíma o to, kam smeruje stavebníctvo vo veku dát. Artem neškriabe len po povrchu, ale ponára sa hlboko do súčasného vývoja, výziev a sľubných príležitostí v stavebníctve. Čo túto knihu odlišuje, je jej prístupnosť - Artem vysvetluje zložité myšlienky pomocou zrozumiteľných analógií, ktoré uľahčujú pochopenie obsahu. Kniha je pre mňa neuveriteľne informatívna a zároveň pútavá. Celkovo možno povedať, že Artem vytvoril cenný zdroj informácií, ktorý nielen informuje, ale aj inšpiruje. Či už ste skúsený profesionál, alebo nováčik v stavebníctve, táto kniha rozšíri váš rozhlás a prehľbi vaše chápanie toho, kam toto odvetvie smeruje. Vrelo odporúčame!"

- **Moayad Saleh**, architekt a manažér implementácie BIM v spoločnosti TMM GROUP Gesamtplanungs GmbH.

"Musím povedať, že kniha Data-Driven Construction je hodná toho, aby sa vyučovala ako učebnica na univerzitách, a je to kniha, ktorá bude cenným prínosom pre rozvíjajúcu sa oblasť BIM. Kniha Data-Driven Construction obsahuje technický slovník, ktorý veľmi dobre vysvetluje pojmy. Témky, ktoré sa mimoriadne ľažko vysvetľujú, sú zjednodušené a zrozumiteľné veľmi pekným vizuálnym jazykom. Myslím si, že to, čo má byť vysvetlené na vizuálnej stránke, by malo byť čitateľovi vyjadrené, aj keď stručne. Zrozumiteľnosť niektorých vizuálov, inými slovami, čítanie vizuálu si vyžaduje samostatné informácie. Chcel by som tiež povedať, že hodnotnú prácu Artema Boika rád predstavím na svojich prednáškach a seminároch na univerzitách."

- **Dr. Ediz Yazicioğlu**, majiteľ spoločnosti ArchCube a prednášajúci na Katedre architektúry Istanbulskej technickej univerzity a na Univerzite Medipol v odbore riadenia stavebných projektov.

"Data Driven Construction názorne sprostredkúva základy práce s údajmi o budovách na základe informácií. Kniha, ktorá sa zaobrá informačným tokmi a základnými ekonomickými pojmom, a tým sa odlišuje od ostatných kníh o BIM, pretože predstavuje nielen pohľad výrobca softvéru, ale snaží sa sprostredkovať aj základné pojmy. Kniha, ktorá sa oplatí prečítať a vidieť."

- **Jakob Hirn**, generálny riaditeľ a spoluzakladateľ spoločnosti Build Informed GmbH a iniciátor inovačného fóra "On Top With BIM".

"Údaje sú nová ropa", ako sa hovorí, takže ich vyhľadávači alebo tažiči by mali mať správne nástroje a myšlenie, aby z tohto zdroja 21. storočia získali hodnotu. Stavebníctvo sa už príliš dlho pohybuje na šíkmej ploche procesov založených na "3D informáciách", pričom dodávka projektu je založená na niekym upečených informáciách (napr. už vykreslili koláčový alebo stĺpcový graf), zatiaľ čo podkladové "dáta" (napr. nespracovaná tabuľka) sú schopné priniesť oveľa viac, najmä preto, že fúzia viacerých údajov a umelá inteligencia prinášajú neobmedzený potenciál. Ak dodávate (alebo vyučujete/vykonávate výskum) konštrukcie, táto kniha je vaším najlepším - a zatiaľ jediným - zdrojom informácií na orientáciu vo svete založenom na údajoch, v ktorom sme sa ocitli."

- **Dr. Zulfikar Adamu**, docent strategického IT v stavebníctve na LSBU, Veľká Británia.

"Data-Driven Construction" od Artema Boika je pôsobivé dielo, ktoré ponúka solídný základ pre stavebníctvo v čase neustále rastúcich technológií a informačných možností. Boikovi sa darí zrozumiteľne prezentovať zložité témy a zároveň predstaviť viacionské myšlienky. Kniha je premysleným kompendiom, ktoré nielen poukazuje na súčasný vývoj, ale poskytuje aj výhľad na budúce inovácie. Odporúčame ju každému, kto chce preniknúť do problematiky plánovania a realizácie stavieb na základe údajov."

- **Markus Eiberger**, prednášajúci na Univerzite aplikovaných vied v Stuttgartre, hlavný projektový manažér a zástupca vedúceho pobočky v spoločnosti Konstruktionsgruppe Bauen, člen predstavenstva združenia BIM Cluster Baden-Württemberg.



Druhé vydanie, marec 2025.  
© 2025 | Artem Boiko | Karlsruhe

ISBN 978-3-912002-17-1



Artem Boiko Autorské práva

[boikoartem@gmail.com](mailto:boikoartem@gmail.com)  
[info@datadrivenconstruction.io](mailto:info@datadrivenconstruction.io)

Žiadna časť tejto knihy nesmie byť reprodukovaná alebo prenášaná v akejkoľvek forme alebo akýmkoľvek prostriedkami, elektronickými alebo mechanickými, vrátane fotokopírovania, nahrávania alebo akéhokoľvek systému na ukladanie a vyhľadávanie informácií, bez písomného súhlasu autora - s výnimkou nekomerčného šírenia v nezmenenej podobe. Kniha sa šíri bezplatne a môže sa voľne poskytovať iným používateľom na osobné, vzdelenacie alebo výskumné účely za predpokladu, že sa zachová autorstvo a odkazy na originál. Autor si ponecháva všetky nevlastnícke práva na text a neposkytuje žiadne výslovné ani implicitné záruky. Spoločnosti, výrobky a názvy uvedené v knihe môžu byť fiktívne alebo použité ako príklady. Autor nezodpovedá za žiadne dôsledky použitia uvedených informácií. Informácie obsiahnuté v knihe sa poskytujú "tak, ako sú", bez záruky úplnosti alebo relevantnosti. Autor nezodpovedá za náhodné alebo následné škody, ktoré vzniknú v dôsledku používania informácií, kódu alebo programov obsiahnutých v tejto knihe. Ukážky kódu uvedené v tejto knihe slúžia len na vzdelenacie účely. Čitatelia ich používajú na vlastné riziko. Autor odporúča, aby sa všetky softvérové riešenia pred použitím v produkčnom prostredí otestovali. Všetky ochranné známky a názvy produktov uvedené v texte sú ochranné známky, registrované ochranné známky alebo servisné značky príslušných spoločností a sú majetkom ich príslušných vlastníkov. Použitie týchto názvov v knihe neznamená žiadny vzťah s ich vlastníkmi ani ich podporu. Zmienka o produktoch alebo službách tretích strán nepredstavuje odporúčanie ani neznamená ich schválenie. Názvy spoločností a produktov použité v príkladoch môžu byť ochrannými známkami ich vlastníkov. Odkazy na webové stránky tretích strán sa uvádzajú ako pomôcka a neznamenajú, že autor schvaľuje informácie uvedené na týchto stránkach. Všetky uvedené štatistiky, citácie a štúdie boli aktuálne v čase písania článku. Údaje sa môžu časom meniť.

Táto kniha je šírená pod licenciou Creative Commons Uznanie autorstva - Nekomerčné - Bez odvodení 4.0 International (CC BY-NC-ND 4.0). Môžete ju kopírovať a šíriť na nekomerčné účely za predpokladu, že si zachováte autorstvo a nebudeťe robiť žiadne zmeny.



CC BY-NC-ND 4.0

© 2024 Artyom Boiko. Prvé vydanie.  
© 2025 Artyom Boiko. Druhé, prepracované a doplnené vydanie.  
Všetky práva vyhradené.

## PREDSLOV K DRUHÉMU VYDANIU

Táto kniha je výsledkom živého dialógu s odbornou verejnosťou. Vychádza z mnohých odborných diskusií o správe údajov v stavebnictve, ktoré sa uskutočnili na rôznych odborných platformách a sociálnych sieťach. Tieto diskusie sa stali základom pre články, publikácie a obrazové materiály, ktoré vyvolali širokú odozvu v odbornej komuniti. Autorský obsah každoročne pritahuje milióny zobrazení na rôznych platformách a v rôznych jazykoch a spája odborníkov v oblasti digitalizácie stavebnictva.

Do roka od vydania prvého vydania si knihu objednali odborníci z viac ako 50 krajín, od Brazílie a Peru až po Maurícius a Japonsko. Druhé vydanie knihy, ktoré práve držíte v rukách, bolo prepracované a rozšírené na základe spätej väzby odborníkov, kritiky prvého vydania a diskusií v odborných kruhoch. Vďaka spätej väzbe bolo druhé vydanie výrazne rozšírené: pribudli nové kapitoly o technológiách CAD (BIM) a o vytváraní efektívnych ETL -procesov. Výrazne sa zvýšil aj počet praktických príkladov a prípadových štúdií. Mimoriadne cenná je spätná väzba od vedúcich predstaviteľov stavebného priemyslu, konzultačných spoločností a veľkých IT-spoločností, ktorí sa na autora obrátili s otázkami digitalizácie a interoperability pred vydaním prvej verzie knihy aj po ňom. Mnohí z nich už uplatnili prístupy opísané v knihe alebo ich plánujú uplatniť v blízkej budúcnosti.

Držíte v rukách knihu, ktorá vznikla na základe diskusie a aktívnej výmeny názorov. Pokrok sa rodí v dialógu, v strete názorov a otvorenosti novým prístupom. Ďakujeme, že ste sa stali súčasťou tohto dialógu. Vaša konštruktívna kritika je základom budúcich zlepšení. Ak v texte nájdete chyby alebo sa chcete podeliť o nápady a návrhy, každá spätná väzba je vítaná. Kontaktné údaje sú uvedené na konci knihy.

## PREČO JE KNIHA ZADARMO?

Táto kniha bola koncipovaná ako otvorený vzdelávací zdroj zameraný na šírenie moderných prístupov k správe údajov v stavebnictve. Prvá verzia knihy slúžila ako základ pre zber pripomienok a návrhov odbornej komunity, ktoré umožnili zlepšiť štruktúru a obsah materiálu. Všetky pripomienky, návrhy a nápady boli starostlivo analyzované a zapracované do tejto revidovanej verzie. Cieľom knihy je pomôcť odborníkom v stavebnictve pochopiť význam práce s údajmi: systematicky, vedome a s ohľadom na dlhodobú hodnotu informácií. Autor zozbieral príklady, ilustrácie a praktické postrehy z viac ako 10 rokov práce v oblasti digitalizácie stavebnictva. Väčšina tohto materiálu vznikla na základe reálnych projektov, diskusií s inžiniermi a developermi, účasti na medzinárodných iniciatívach a vzdelávacích seminároch. Kniha je pokusom o štruktúrovanie nahromadených skúseností a ich zdieľanie prístupným spôsobom. Ak chcete podporiť ďalšie šírenie myšlienok knihy a získať pohodlný formát na čítanie, prácu s príkladmi a vizuálnymi materiálmi - môžete si zakúpiť [tlačenú verziu](#).

## PRÁVA NA POUŽÍVANIE

Všetky materiály, ilustrácie a fragmenty tejto knihy sa môžu reprodukovať, citovať alebo používať v akomkoľvek formáte a na akomkoľvek médiu pod podmienkou, že sa uvedie zdroj: autorstvo Artem Boiko a názov knihy "Data-Driven Construction". Ďakujeme za úctu k práci a šíreniu poznatkov.

S úprimnou vdăkou venujem túto knihu svojej rodine, ktorá mi od  
útleho veku vštepovala hlbokú lásku k staviteľstvu, svojmu  
rodnému baníckemu mestu za lekcie odolnosti a svojej manželke  
geodetke, ktorej neochvejná podpora bola mojom neustálou  
inšpiráciou.

## **PRE KOHO JE TÁTO KNIHA URČENÁ**

Táto kniha, napísaná prístupným jazykom, je určená širokému okruhu čitateľov v stavebníctve - od študentov a začiatočníkov, ktorí chcú pochopiť základy moderných stavebných procesov, až po odborníkov, ktorí potrebujú aktuálnu metodiku na správu údajov v stavebníctve. Či už ste architekt, inžinier, majster, stavebný manažér alebo dátový analytik, táto komplexná príručka s množstvom jedinečných ilustrácií a grafov ponúka cenné poznatky o tom, ako využívať dátá v podnikaní na optimalizáciu a automatizáciu procesov, zlepšenie rozhodovania a riadenie stavebných projektov na rôznych úrovniach pomocou moderných nástrojov.

Kniha je komplexnou príručkou, ktorá spája teoretické základy a praktické odporúčania na integráciu techník správy údajov do stavebných procesov. Kniha sa zameriava na strategické využívanie informácií na optimalizáciu činností, automatizáciu procesov, zlepšenie rozhodovania a efektívne riadenie projektov pomocou moderných digitálnych nástrojov.

Táto kniha zahŕňa teoretické a praktické aspekty práce s informáciami v stavebníctve. Prostredníctvom podrobnychých príkladov skúma metodiku parametrizácie úloh, zberu požiadaviek, spracovania neštruktúrovaných a multiformátových údajov a ich transformácie do efektívnych riešení pre stavebné firmy.

Čitateľ postupne prejde cestu od tvorby požiadaviek a vývoja základných dátových modelov až po zložitejšie procesy integrácie heterogénnych informačných zdrojov, tvorby ETL -procesov, konštrukcie informačných potrubí a modelov strojového učenia. Postupný prístup umožňuje prehľadne demonštrovať mechanizmy organizácie a automatizácie podnikových procesov a systémov na podporu rozhodovania v stavebníctve. Každú časť knihy uzatvára praktická kapitola obsahujúca návody krok za krokom, ktoré umožňujú okamžitú aplikáciu získaných poznatkov v reálnych projektoch.

## ZHRNUTIE ČASTÍ KNIHY

Táto kniha je štruktúrovaná okolo koncepcie transformácie údajov v hodnotovom reťazci: od zberu údajov a zabezpečenia ich kvality až po analytické spracovanie a získanie hodnotných praktických riešení pomocou moderných nástrojov a metodík.

**Časť 1: Digitálny vývoj v stavebníctve** - sleduje historickú transformáciu správy údajov od hlinených tabuliek k moderným digitálnym systémom, analyzuje vznik modulárnych systémov a rastúci význam digitalizácie informácií v kontexte priemyselných revolúcií.

**Časť 2: Informačné výzvy pre stavebníctvo** - skúma problémy fragmentácie údajov, "informačných síl", vplyv prístupu HiPPO rozhodovanie a obmedzenia proprietárnych formátov, pričom navrhuje zvážiť prechod na ekosystémy umelej inteligencie a LLM.

**Časť 3: Systematizácia údajov v stavebníctve** - vytvára typológiu údajov v stavebníctve, opisuje metódy ich organizácie, integrácie s podnikovými systémami a zaoberá sa vytváraním kompetenčných centier pre štandardizáciu informačných procesov.

**Časť 4: Zabezpečenie kvality údajov** - odhaluje metodiky na premenu rôznorodých informácií na kvalitné štruktúrované údaje vrátane extrakcie údajov z rôznych zdrojov, validácie a modelovania pomocou LLM.

**Časť 5: Výpočty nákladov a času** - zaoberá sa digitalizáciou výpočtov nákladov a plánovania, automatizáciou získavania objemov z modelov CAD (BIM), technológiami modelovania 4D-8D a výpočtom ESG stavebných projektov.

**Časť 6: CAD a BIM** - kriticky analyzuje vývoj technológií navrhovania, otázky interoperability systémov, trendy smerom k otvoreným dátovým formátom a perspektívy využitia umelej inteligencie v navrhovaní.

**Časť 7: Analýza údajov a automatizácia** - venuje sa princípom vizualizácie informácií, kľúčovým ukazovateľom výkonnosti, procesom ETL, nástrojom na orchestráciu pracovných tokov a použitiu jazykových modelov na automatizáciu rutinných úloh.

**Časť 8: Ukladanie a správa údajov** - skúma formáty ukladania údajov, koncepcie dátového skladu a dátového jazera, zásady správy údajov a nové prístupy vrátane vektorových databáz a metodík DataOps a VectorOps.

**Časť 9: Big Data a strojové učenie** - zameriava sa na prechod na objektívnu analýzu založenú na historických údajoch, internet vecí na staveniskách a použitie algoritmov strojového učenia na predpovedanie nákladov a harmonogramov projektov.

**Časť 10: Stavebníctvo vo veku digitálnych dát** - predstavuje pohľad na budúcnosť stavebníctva, analyzuje prechod od kauzálnej analýzy k práci s koreláciemi, koncept "uberizácie" stavebníctva a stratégie digitálnej transformácie.

# What is meant by **data-driven construction** ?



# ÚVOD

Ako dlho si vaša spoločnosť dokáže udržať konkurencieschopnosť vo svete, kde sa technológie rýchlo vyvíjajú a každý aspekt podnikania, od časového plánovania a kalkulácie nákladov až po analýzu rizík, sa automatizuje pomocou modelov strojového učenia?

Stavebnictvo, ktoré existuje tak dlho ako ľudstvo samo, je na prahu revolučných zmien, ktoré sľubujú, že úplne zmenia naše predstavy o tradičnom stavebnictve. Už aj v iných odvetviach hospodárstva digitalizácia nielenže mení pravidlá, ale nemilosrdne vytláča z trhu spoločnosti, ktoré sa nedokázali prispôsobiť novému prostrediu spracovania dát a nie sú schopné zvýšiť rýchlosť rozhodovania (obr. 1).



Obr. 1 Rýchlosť rozhodovania v stavebnictve závisí od ľudského faktora častejšie ako v iných odvetviach.

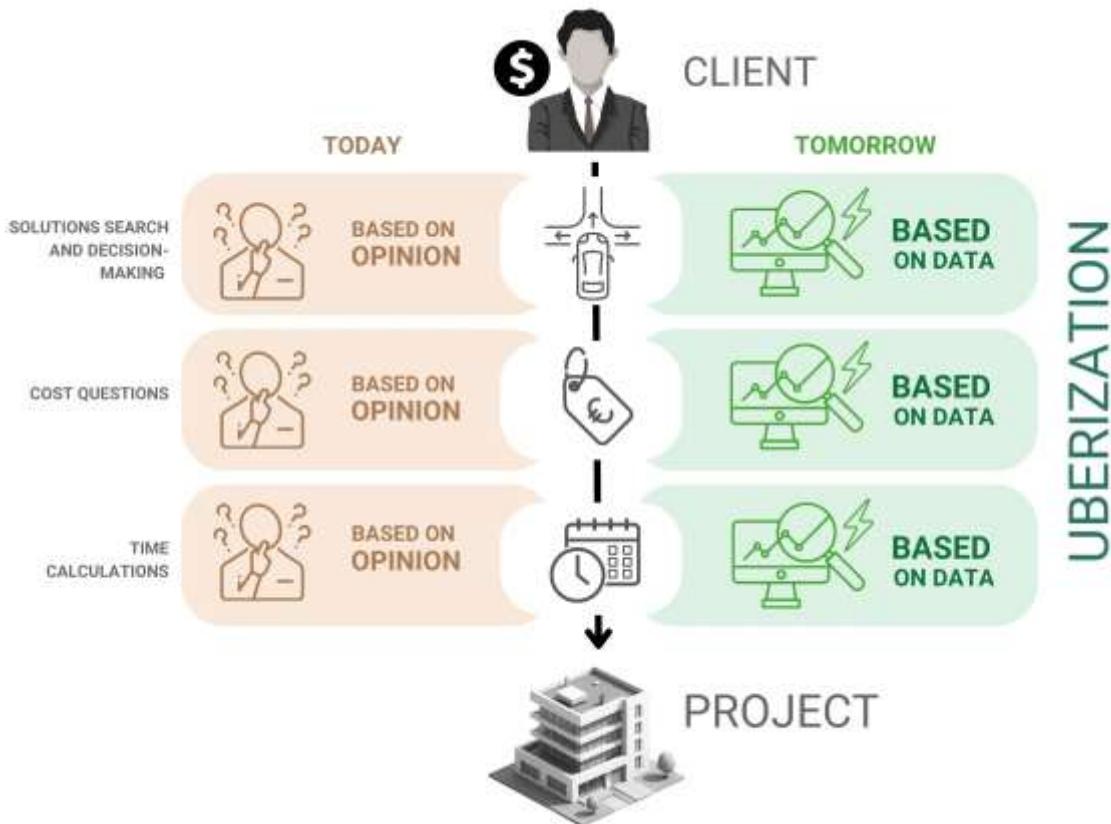
Bankovníctvo, maloobchod, logistika a poľnohospodárstvo rýchlo smerujú k úplnej digitalizácii, kde už nepresnosti a subjektívne názory nemajú miesto. Moderné algoritmy dokážu analyzovať obrovské množstvo údajov a poskytovať zákazníkom presné predpovede - či už ide o pravdepodobnosť splatenia úveru, optimálne trasy dodávok alebo predpovedanie rizík.

Stavebnictvo je jedným z posledných odvetví, ktoré nevyhnutne prechádza od riešení založených na názoroch vysoko platených odborníkov k riešeniam založeným na údajoch. Tento prechod je poháňaný nielen novými technologickými možnosťami, ale aj zvýšenými požiadavkami trhu a zákazníkov na transparentnosť, presnosť a rýchlosť.

Robotizácia, automatizácia procesov, otvorené údaje a na nich založené prognózy - to všetko už nie sú len možnosti, ale nevyhnutnosti. Väčšina spoločností v stavebnictve, ktoré boli ešte nedávno zodpovedné voči klientovi za kalkuláciu objemu, nákladov, času projektov a kontrolu kvality, teraz

riskuje, že sa zmenia na obyčajných vykonávateľov zákaziek, ktorí nebudú robiť kľúčové rozhodnutia (obr. 2).

Vďaka pokroku vo výpočtovom výkone, algoritmom strojového učenia a demokratizácii prístupu k údajom je teraz možné automaticky kombinovať údaje z rôznych zdrojov, čo umožňuje hlbšiu analýzu procesov, predpovedanie rizík a optimalizáciu nákladov v diskusných fázach stavebného projektu. Tieto technológie majú potenciál radikálne zvýšiť efektívnosť a znížiť náklady v celom odvetví.



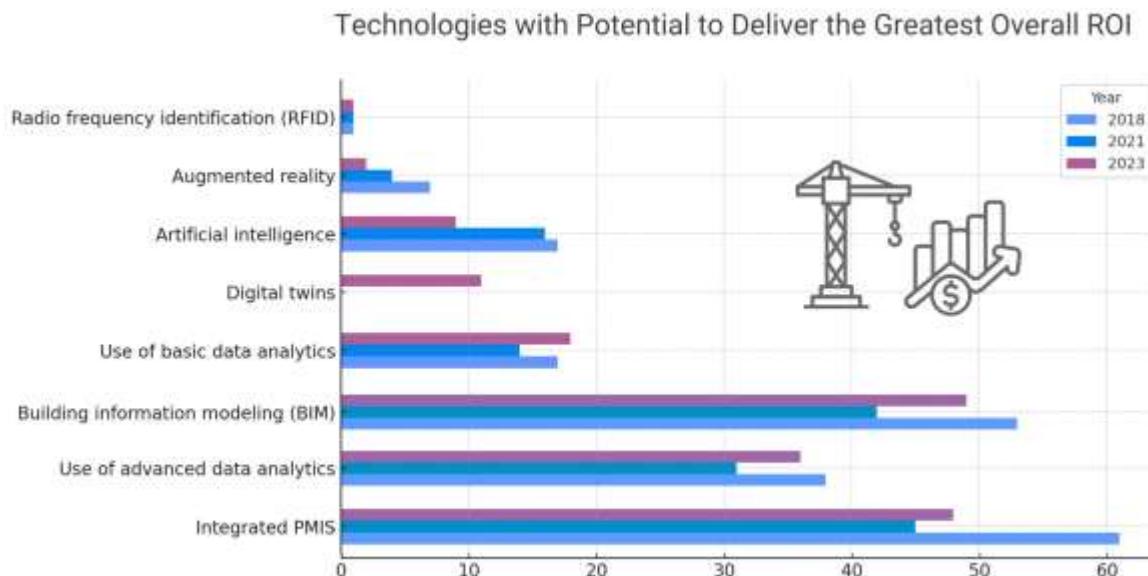
Obr. 2 Klient nemá záujem o nadmerný ľudský faktor na ceste k realizácii svojho projektu.

Napriek všetkým výhodám nových nástrojov a koncepcii stavebníctva v zavádzaní nových technológií výrazne zaostáva za ostatnými odvetviami hospodárstva.

Podľa správy IT Metrics Key Data 2017 je stavebníctvo na poslednom mieste vo výdavkoch na IT spomedzi 19 ďalších odvetví [1].

Rýchly nárast objemu údajov a zložitosť procesov sa stáva pre vedenie spoločnosti bolestivým problémom a hlavným problémom pri používaní nových technológií je, že údaje napriek ich množstvu zostávajú roztrieštené, neštruktúrované a často nekompatibilné medzi rôznymi systémami a softvérovými produktmi. Preto sa mnohé spoločnosti v stavebnom sektore v súčasnosti zaoberajú predovšetkým problémami s kvalitou údajov, ktoré sa dajú vyriešiť len zavedením efektívnych, automatizovaných systémov riadenia a analýzy.

Podľa prieskumu, ktorý uskutočnila spoločnosť KPMG® medzi manažérmi v stavebnictve v roku 2023 [2], majú najväčší potenciál na zlepšenie návratnosti investícií do projektov informačné systémy na riadenie projektov (PMIS), pokročilá a základná analýza údajov a informačné modelovanie budov (BIM) (obrázok 3).



Obr. 3 Prieskum medzi manažérmi stavebných spoločností: Ktoré technológie zabezpečia najvyššiu návratnosť investícií (ROI) do investičných projektov? (na základe materiálov [2]).

Riešením problémov spojených s integráciou údajov do podnikových procesov je zabezpečenie vysokej kvality informácií, používanie vhodných formátov údajov a uplatňovanie účinných metód na vytváranie, ukladanie, analýzu a spracovanie údajov.

Uvedomenie si hodnoty údajov nútí rôzne priemyselné odvetvia upustiť od izolovaných aplikácií a zložitých byrokratických štruktúr riadenia. Namiesto toho sa pozornosť presúva na vytváranie nových prístupov k informačnej architektúre, čím sa spoločnosti menia na moderné podniky založené na údajoch. Skôr či neskôr tento krok urobí aj samotné stavebnictvo, ktoré prejde od postupnej digitálnej evolúcie k skutočnej digitálnej revolúcii, ktorá sa dotkne všetkých spoločností.

Prechod na obchodné procesy založené na údajoch nebude jednoduchý. Mnohé spoločnosti budú celiť výzvam, pretože vedúci pracovníci nie vždy rozumejú tomu, ako využiť chaotické súbory údajov na zlepšenie efektívnosti a rastu podniku.

Táto kniha sa zaobrá svetom údajov, v ktorom sa informácie stávajú kľúčovým strategickým zdrojom, ktorý určuje efektívnosť a udržateľnosť podnikových procesov. S rýchlym rastom informácií čelia spoločnosti novým výzvam. Digitálna transformácia už nie je len módnym slovom - stáva sa nevyhnutnosťou.



Obr. 4 Základom konštrukcie sú údaje a procesy.

Pochopiť transformáciu znamená vedieť vysvetliť zložité jednoduchými slovami. Preto je kniha napísaná prístupným jazykom a sprevádzajú ju autorove ilustrácie, vytvorené špeciálne na jasné vysvetlenie kľúčových pojmov. Tieto diagramy, grafy a vizualizácie sú navrhnuté tak, aby odstránili bariéry vnímania a urobili materiál zrozumiteľný aj pre tých, ktorí predtým považovali takéto témy za príliš zložité. Všetky ilustrácie, diagramy a grafy v tejto knihe sú vytvorené autorom a sú navrhnuté špeciálne na vizualizáciu kľúčových pojmov opísaných v texte.

*Jeden obrázok vydá za tisíc slov [3].*

- Fred R. Barnard, anglický ilustrátor, 1927.

Na prepojenie teórie s praxou budeme používať nástroje umelej inteligencie (najmä jazykové modely), ktoré vám umožnia vyvíjať riešenia bez potreby hlbokých znalostí programovania. Ak ste orientovaní na praktický materiál a viac vás zaujíma praktická práca s údajmi, môžete preskočiť prvú úvodnú časť a prejsť rovno na druhú časť knihy, kde sa začína opis konkrétnych príkladov a prípadov.

Nevkladajte však príliš veľké očakávania do nástrojov umelej inteligencie, strojového učenia a LLM (Large Language Models) vo všeobecnosti. Bez kvalitných vstupných údajov a hlbokého pochopenia predmetu ani tie najpokročilejšie algoritmy nemôžu poskytnúť spoľahlivé a zmysluplné výsledky.

Generálny riaditeľ spoločnosti Microsoft Satya Nadella varuje pred rizikom bubliny v oblasti umelej inteligencie na začiatku roka 2025 [4] a prirovnáva súčasný hype k bubline dot-com. Zdôrazňuje, že tvrdenia o dosiahnutí míľníkov AGI (Artificial General Intelligence) bez náležitého zdôvodnenia sú "nezmyselnou manipuláciou s metrikami". Nadella sa domnieva, že skutočný úspech AI by sa mal merať podľa jej príspevku k rastu globálneho HDP, a nie podľa prílišného zdôrazňovania módnych slov.

Za všetkými slovami o nových technológiách a koncepciách sa skrýva zložitá a náročná práca na

zabezpečenie kvality údajov, parametrizácia podnikových procesov a prispôsobenie nástrojov reálnym úlohám.

Prístup založený na údajoch nie je produkt, ktorý si môžete jednoducho stiahnuť alebo kúpiť. Je to stratégia, ktorú treba vybudovať. Začína sa novým pohľadom na existujúce procesy a problémy a potom si vyžaduje disciplinovaný pohyb zvoleným smerom.

Poprední vývojári softvéru a dodávateelia aplikácií nebudú hnacím motorom zmien v stavebníctve, pre mnohých z nich je prístup založený na údajoch hrozou pre ich zavedený obchodný model.

*Iné odvetvia [na rozdiel od stavebníctva], ako napríklad automobilový priemysel, už prešli radikálnou a prevratnou zmenou a ich digitálna transformácia je v plnom prúde. Stavebné spoločnosti musia konáť rýchlo a rozhodne: pohotové spoločnosti budú žať obrovské zisky, zatiaľ čo pre tie, ktoré zaváhajú, budú riziká vážne. Spomeňte si na prevrat, ktorý v tomto odvetví spôsobila digitálna fotografia [5].*

- Správa Svetového ekonomického fóra Formovanie budúcnosti stavebníctva, 2016.

Spoločnosti, ktoré včas rozpoznajú možnosti a výhody nového prístupu, získajú udržateľnú konkurenčnú výhodu a budú sa môcť rozvíjať a rásť bez závislosti od riešení od veľkých dodávateľov.

Toto je vaša šanca nielen prežiť nadchádzajúcemu búrku digitalizácie informácií, ale aj prevziať nad ňou kontrolu. V tejto knihe nájdete nielen analýzu súčasného stavu odvetvia, ale aj konkrétnie odporúčania na prehodnotenie a reštrukturalizáciu vašich procesov a vášho podnikania, aby ste sa stali lídom v novej ére stavebníctva a zvýšili svoje profesionálne skúsenosti.

Digitálna budúcnosť stavebníctva nie je len o využívaní nových technológií a programov, ale aj o zásadnom prehodnotení spracovania údajov a obchodných modelov.

Je vaša spoločnosť pripravená na túto strategickú zmenu?



## OBSAH

<b>ÚVOD .....</b>	<b>1</b>
<b>OBSAH.....</b>	<b>I</b>
<b>II ČASŤ OD HLINENÝCH TABULIEK PO DIGITÁLNU REVOLÚCIU: AKO SA VYVÍJALI INFORMÁCIE V STAVEBNÍCTVE .....</b>	<b>2</b>
KAPITOLA 2.1. VÝVOJ POUŽÍVANIA ÚDAJOV V STAVEBNÍCTVE.....	3
Zrod dátovej éry v stavebníctve .....	3
Od hliny a papyrusu k digitálnej technológií.....	4
Proces ako nástroj pre skúsenosti založené na údajoch.....	5
Digitalizácia informácií o stavebnom procese.....	7
KAPITOLA 2.2. TECHNOLÓGIE A SYSTÉMY RIADENIA V MODERNOM STAVEBNÍCTVE .....	11
Digitálna revolúcia a vznik modulárnych MRP/ERP -systémov .....	11
Systémy správy údajov: od dolovania údajov k obchodným výzvam.....	13
Firemné mycélium: ako sa údaje spájajú s podnikovými procesmi .....	17
KAPITOLA 2.3. DIGITÁLNA REVOLÚCIA A EXPLÓZIA DÁT .....	20
Začiatok dátového boomu ako evolučná vlna .....	20
Množstvo údajov generovaných v modernej spoločnosti.....	22
Náklady na ukladanie údajov: ekonomický aspekt .....	23
Hranice akumulácie údajov: od hmotnosti k významu.....	25
Ďalšie kroky: od teórie údajov k praktickej zmene.....	27
<b>III ČASŤ AKO SA STAVEBNÍCTVO TOPÍ V DÁTOVOM CHAOSE.....</b>	<b>28</b>
KAPITOLA 3.1. ROZTRIEŠTENOSŤ ÚDAJOV A DÁTOVÉ SILÁ.....	29
Čím viac nástrojov, tým efektívnejšie podnikanie? .....	29
Dátové silá a ich vplyv na výkonnosť spoločnosti .....	31
Duplicita a nedostatočná kvalita údajov ako dôsledok nejednotnosti.....	34
HiPPO alebo nebezpečenstvo názorov pri rozhodovaní .....	36
Neustály nárast zložitosti a dynamiky obchodných procesov.....	39
Štvrtá priemyselná revolúcia (Priemysel 4.0) a piata priemyselná revolúcia (Priemysel 5.0) v stavebníctve .....	42
KAPITOLA 3.2. PREMENA CHAOSU NA PORIADOK A ZNÍŽENIE ZLOŽITOSTI .....	45
Zbytočný kód a uzavreté systémy ako prekážka zvyšovania produktivity.....	45
Od silosov k jednotnému dátovému skladu .....	47

Integrované úložné systémy umožňujú prechod na agentov s umelou inteligenciou .....	48
Od zberu údajov k rozhodovaniu: cesta k automatizácii.....	51
Ďalšie kroky: premena chaosu na riadený systém .....	53
<b>IV ČASŤ DÁTOVÝ RÁMEC V OBCHODNÝCH PROCESOCH V STAVEBNÍCTVE .....55</b>	
KAPITOLA 4.1. DÁTOVÉ TYPY V STAVEBNÍCTVE .....	56
Najdôležitejšie typy údajov v stavebníctve.....	56
Štruktúrované údaje .....	60
Relačné databázy RDBMS a dotazovací jazyk SQL .....	61
SQL - dopyty v databázach a nové trendy .....	64
Neštruktúrované údaje.....	66
Textové údaje: medzi neštruktúrovaným chaosom a štruktúrovaným chaosom y .....	67
Čiastočne štruktúrované a voľne štruktúrované údaje.....	68
Geometrické údaje a ich použitie.....	69
Údaje CAD: od návrhu po ukladanie údajov .....	72
Vznik koncepcie BIM (BOM) a využívanie CAD v procesoch .....	75
KAPITOLA 4.2. HARMONIZÁCIA A ŠTRUKTÚROVANIE ÚDAJOV .....	81
Napĺňanie systémov údajmi v stavebníctve.....	81
Transformácia údajov: rozhodujúci základ modernej podnikovej analýzy .....	83
Dátové modely: vzťahy v údajoch a vzťahy medzi prvkami.....	87
Proprietárne formáty a ich vplyv na digitálne procesy .....	91
Otvorené formáty menia prístup k digitalizácii .....	95
Zmena paradigmy: Open Source ako koniec éry dominancie dodávateľov softvéru.....	96
Štruktúrované otvorené údaje: základ digitálnej transformácie .....	99
KAPITOLA 4.3. LLM A ICH ÚLOHA PRI SPRACOVANÍ ÚDAJOV A OBCHODNÝCH PROCESOCH .....	102
Chatovacie miestnosti LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok na automatizáciu dátových procesov.....	102
Veľké jazykové modely LLM: ako to funguje .....	103
Používanie miestnych LLM pre citlivé firemné údaje .....	106
Úplná kontrola nad AI vo firme a spôsob nasadenia vlastného LLM .....	108
RAG: Inteligentný LLM -asistenti s prístupom k podnikovým údajom.....	110
KAPITOLA 4.4. IDE S PODPOROU LLM A BUDÚCE PROGRAMOVÉ ZMENY .....	112
Výber IDE: od experimentov LLM po obchodné riešenia .....	112
IDE s podporou LLM a budúce zmeny v programovaní.....	114
Python Pandas: nenahraditeľný nástroj na prácu s údajmi .....	115

DataFrame: univerzálny formát tabuľkových údajov .....	119
Ďalšie kroky: budovanie udržateľného dátového rámca .....	122
<b>V ČASŤ KVALITA ÚDAJOV: ORGANIZÁCIA, ŠTRUKTÚROVANIE, MODELOVANIE ..... 124</b>	
KAPITOLA 5.1. PREVOD ÚDAJOV DO ŠTRUKTÚROVANEJ PODOBY ..... 125	
Naučte sa, ako meniť dokumenty, súbory PDF, obrázky a texty do štruktúrovaných formátov. ....	125
Príklad konverzie PDF -dokumentu na tabuľku .....	126
Konvertovať obrázok JPEG, PNG do štruktúrovanej podoby .....	130
Konverzia textových údajov do štruktúrovanej podoby.....	133
Konverzia údajov CAD (BIM) do štruktúrovanej podoby .....	136
Dodávatelia riešení CAD prechádzajú na štruktúrované údaje.....	141
KAPITOLA 5.2. KLASIFIKÁCIA A INTEGRÁCIA: SPOLOČNÝ JAZYK PRE STAVEBNÉ ÚDAJE....144	
Rýchlosť rozhodovania závisí od kvality údajov .....	144
Štandardizácia a integrácia údajov.....	145
Digitálna interoperabilita sa začína požiadavkami .....	148
Spoločný jazyk konštrukcie: úloha klasifikátorov v digitálnej transformácii .....	150
Masterformat, OmniClass, Uniclass a CoClass: vývoj klasifikačných systémov .....	153
KAPITOLA 5.3. MODELOVANIE ÚDAJOV A CENTRUM EXCELENTNOSTI .....	158
Modelovanie údajov: koncepcný, logický a fyzický model.....	158
Praktické modelovanie údajov v kontexte stavebníctva .....	161
Vytvorenie databázy pomocou LLM .....	164
Centrum excelentnosti (CoE) pre modelovanie údajov .....	166
KAPITOLA 5.4. SYSTEMATIZÁCIA POŽIADAVIEK A OVEROVANIE INFORMÁCIÍ .....	169
Zhrubažďovanie a analýza požiadaviek: transformácia komunikácie na štruktúrované údaje.....	169
Procesné diagramy a účinnosť koncepcných rámcov .....	173
Štruktúrované požiadavky a regulárne výrazy RegEx .....	175
Zber údajov pre proces overovania.....	180
Overenie údajov na stránke a výsledky overenia .....	182
Vizualizácia výsledkov overovania.....	187
Porovnanie kontrol kvality údajov s potrebami ľudského života .....	189
Ďalšie kroky: premena údajov na presné výpočty a plány.....	191
<b>VI ČASŤ VÝPOČTY NÁKLADOV A ČASU: ZAČLENENIE ÚDAJOV DO STAVEBNÝCH PROCESOV .....</b>	<b>193</b>
KAPITOLA 6.1. VÝPOČTY A ODHADY NÁKLADOV NA STAVEBNÉ PROJEKTY.....	194

---

Základy stavebníctva: odhad množstva, nákladov a času .....	194
Metódy výpočtu odhadovaných nákladov na projekty .....	195
Metóda založená na zdrojoch kalkulácia nákladov a odhad v stavebníctve .....	196
Databáza stavebných zdrojov: katalóg stavebných materiálov a prác .....	196
Zostavenie výpočtov a vyčíslenie nákladov na práce na základe zdrojov .....	198
Konečné vyčíslenie nákladov na projekt: od odhadov k rozpočtom .....	203
<b>KAPITOLA 6.2. VÝPOČET MNOŽSTVA A AUTOMATICKE VYTVÁRANIE ODHADOV A HARMONOGRAMOV .....</b>	<b>207</b>
Prechod z 3D na 4D a 5D: využitie objemových a kvantitatívnych parametrov .....	207
5D atribúty a získavanie objemov atribútov z CAD .....	207
QTO Quantity Take-Off: zoskupovanie údajov o projekte podľa atribútov .....	211
Automatizácia QTO pomocou LLM a štruktúrovaných údajov .....	216
QTO výpočet celého projektu pomocou skupinových pravidiel z tabuľky Excel .....	220
<b>KAPITOLA 6.3. 4D, 6D -8D A VÝPOČET EMISIÍ OXIDU UHLIČITÉHO CO<sub>2</sub>.....</b>	<b>226</b>
4D model: integrácia času do stavebných odhadov .....	226
Harmonogram výstavby a jeho automatizácia na základe údajov o nákladoch .....	227
Rozšírené vrstvy atribútov 6D -8D: od energetickej účinnosti po zaistenie bezpečnosti .....	229
Odhad emisií CO <sub>2</sub> a výpočet emisií oxidu uhličitého zo stavebných projektov .....	232
<b>KAPITOLA 6.4. STAVEBNÝ ERP A PMIS SYSTÉMY .....</b>	<b>236</b>
Stavebný ERP -systémy na príklade výpočtov a odhadov .....	236
PMIS: sprostredkovateľ medzi ERP a stavbou .....	241
Špekulácie, zisk, izolovanosť a nedostatok transparentnosti v ERP a PMIS .....	242
Koniec éry uzavretých ERP /PMIS: stavebníctvo potrebuje nové prístupy .....	245
Ďalšie kroky: efektívne využívanie údajov o projektoch .....	246
<b>VII ČASŤ CAD A BIM: MARKETING, REALITA A BUDÚCNOSŤ PROJEKTOVÝCH ÚDAJOV V STAVEBNÍCTVE .....</b>	<b>249</b>
<b>KAPITOLA 7.1. VZNIK KONCEPCIÍ BIM- V STAVEBNÍCTVE .....</b>	<b>250</b>
História vzniku BIM a otvoreného BIM ako marketingových konceptov dodávateľov CAD- 250	250
Realita BIM: namiesto integrovaných databáz - uzavreté modulárne systémy .....	253
Vznik otvoreného formátu IFC v stavebníctve .....	255
Problém formátu IFC v závislosti od geometrického jadra .....	257
Výskyt v konštrukcii témy sémantiky a ontológie .....	260
Prečo sémantické technológie nespĺňajú očakávania v stavebníctve .....	262
<b>KAPITOLA 7.2. UZAVRETÉ FORMÁTY PROJEKTOV A OTÁZKY INTEROPERABILITY .....</b>	<b>266</b>
Uzavreté údaje a klesajúca produktivita: slepá ulička odvetvia CAD (BIM) .....	266

Mýtus o interoperabilite medzi systémami CAD .....	268
Prejsť na stránku USD a granulárne údaje m .....	272
<b>KAPITOLA 7.3. GEOMETRIA V STAVEBNÍCTVE: OD ČIAR PO METRE KUBICKÉ .....</b>	<b>276</b>
Keď sa čiary menia na peniaze alebo prečo stavitelia potrebujú geometriu .....	276
Od čiar k objemom: Ako sa z plochy a objemu stávajú údaje .....	276
Prechod na MESH, USD a polygóny: používanie teselačného zobrazenia pre geometriu ....	278
LOD, LOI, LOMD - jedinečná klasifikácia detailov v CAD (BIM).....	280
Nové normy CAD (BIM) - AIA, BEP, IDS, LOD, COBie.....	283
<b>KAPITOLA 7.4. PARAMETRIZÁCIA NÁVRHU A POUŽITIE LLM PRE CAD PREVÁDZKU.....</b>	<b>288</b>
Ilúzia jedinečnosti údajov CAD (BIM): cesta k analytike a otvoreným formátom .....	288
Návrh prostredníctvom parametrov: budúcnosť CAD a BIM .....	291
Vznik LLM v procesoch spracovania údajov v CAD.....	294
Automatizovaná analýza súborov DWG pomocou LLM a Pandas .....	297
Ďalšie kroky: prechod od uzavretých formátov k otvoreným údajom .....	303
<b>VIII ČASŤ ROZHODOVANIE ZALOŽENÉ NA ÚDAJOCH, ANALYTika, AUTOMATIZÁCIA A STROJOVÉ UČENIE .....</b>	<b>305</b>
<b>KAPITOLA 8.1. ANALÝZA ÚDAJOV A ROZHODOVANIE ZALOŽENÉ NA ÚDAJOCH .....</b>	<b>306</b>
Údaje ako zdroj pri rozhodovaní.....	306
Vizualizácia údajov: kľúč k pochopeniu a rozhodovaniu.....	309
KPI a ROI .....	311
Prístrojové panely a informačné tabule: vizualizácia metrík pre efektívne riadenie .....	313
Analýza údajov a umenie klášť otázky.....	315
<b>KAPITOLA 8.2. TOK ÚDAJOV BEZ MANUÁLNEJ NÁMAHY: PREČO JE POTREBNÝ ETL.....</b>	<b>318</b>
Automatizácia ETL: zníženie nákladov a zrýchlenie spracovania údajov .....	318
ETL Výpis: zber údajov.....	322
ETL Transformácia: aplikácia pravidiel validácie a transformácie.....	325
ETL Load: Vizualizujte výsledky v tabuľkách a grafoch.....	327
ETL Load: Automatické vytváranie dokumentov PDF .....	332
ETL Load: automatické generovanie dokumentov z FPDF .....	333
ETL Načítanie: Hlásenie a načítanie do iných systémov.....	337
ETL s LLM: Vizualizácia údajov z PDF -dokumentov .....	338
<b>KAPITOLA 8.3. AUTOMATICKÝ ETL DOPRAVNÍK (POTRUBIE .....</b>	<b>343</b>
Potrubie: Automatický dopravník ETL údaje .....	343
Potrubie -ETL proces overovania údajov pomocou LLM .....	347

---

Potrubie -ETL: overenie údajov a informácií o prvkoch projektu v CAD (BIM).....	349
<b>KAPITOLA 8.4. ORCHESTRÁCIA ETL A PRACOVNÝCH POSTUPOV: PRAKTICKÉ RIEŠENIA....</b>	<b>355</b>
DAG a Apache Airflow: automatizácia a orchestrácia pracovných postupov .....	355
Apache Airflow: praktická aplikácia automatizáciu ETL.....	356
Apache NiFi na smerovanie a konverziu údajov .....	360
n8n Low-Code, No-Code orchestrácia procesov .....	361
Ďalšie kroky: prechod od manuálnych operácií k riešeniam založeným na analytike .....	364
<b>IX ČASŤ UKLADANIE A SPRÁVA ÚDAJOV V STAVEBNÍCTVE.....</b>	<b>366</b>
<b>KAPITOLA 9.1. DÁTOVÁ INFRAŠTRUKTÚRA: OD FORMÁTOV UKLADANIA PO DIGITÁLNE ÚLOŽISKÁ .....</b>	<b>367</b>
Dátové atómy: základ efektívneho riadenia informácií .....	367
Ukladanie informácií: súbory alebo údaje .....	368
Ukladanie veľkých objemov údajov: Analýza populárnych formátov a ich efektívnosti .....	370
Optimalizácia ukladania pomocou Apache Parquet .....	373
DWH: Dátový sklad dátové sklady .....	375
Data Lake - vývoj ETL na ELT: od tradičného čistenia k flexibilnému spracovaniu.....	377
Architektúra dátového skladu: synergia skladov a dátových jazier.....	379
CDE, PMIS, ERP alebo DWH a dátové jazero.....	381
<b>KAPITOLA 9.2. SPRÁVA DÁTOVÉHO SKLADU A PREVENCIA CHAOSU.....</b>	<b>385</b>
Vektorové databázy a ohraňujúci rámček .....	385
Riadenie údajov), minimalizmus údajov) a dátová bažina). .....	388
DataOps a VectorOps: nové dátové štandardy .....	391
Ďalšie kroky: od chaotického ukladania k štruktúrovanému ukladaniu .....	393
<b>X ČASŤ VEĽKÉ DÁTA, STROJOVÉ UČENIE A PREDPOVEDE.....</b>	<b>395</b>
<b>KAPITOLA 10.1. VEĽKÉ OBJEMY ÚDAJOV A ICH ANALÝZA .....</b>	<b>396</b>
Veľké dátá v stavebníctve: od intuície k predvídateľnosti.....	396
Spochybňovanie uskutočniteľnosti veľkých dát: korelácia, štatistika a výber vzoriek údajov.....	397
Veľké dátá: analýza údajov zo súboru miliónov stavebných povolení v San Franciscu .....	400
Príklad veľkých dát na základe údajov CAD (BIM).....	406
Internet vecí Internet vecí a inteligentné zmluvy.....	410
<b>KAPITOLA 10.2. STROJOVÉ UČENIE A PREDPOVEDE.....</b>	<b>414</b>
Strojové učenie a umelá inteligencia zmenia spôsob, akým tvoríme .....	414
Od subjektívneho hodnotenia k štatistickej prognóze .....	416
Súbor údajov Titanic: Hello World vo svete analytiky data and big data .....	418

Strojové učenie v praxi: od pasažierov Titaniku po riadenie projektov .....	423
Predpovede a prognózy na základe historických údajov .....	428
Kľúčové koncepty strojového učenia .....	430
<b>KAPITOLA 10.3. PREDPOVEDANIE NÁKLADOV A TERMÍNOV POMOCOU STROJOVÉHO UČENIA .....</b>	<b>433</b>
Príklad použitia strojového učenia na zistenie nákladov a časového harmonogramu projektu .....	433
Predpovedanie nákladov a času projektu pomocou lineárnej regresie .....	435
Predpovede nákladov a času projektu pomocou algoritmu K-najbližší sused (k-NN).....	438
Ďalšie kroky: od skladovania k analýze a prognózovaniu .....	442
<b>XI ČASŤ STAVEBNÍCTVO V DIGITÁLNUM VEKU. PRÍLEŽITOSTI A VÝZVY .....</b>	<b>445</b>
<b>KAPITOLA 11.1. STRATÉGIE PREŽITIA: BUDOVANIE KONKURENČNÝCH VÝHOD .....</b>	<b>446</b>
Korelácie namiesto výpočtov: budúcnosť stavebnej analytiky .....	446
Prístup založený na údajoch v stavebníctve: nová úroveň infraštruktúry .....	449
Digitálna kancelária novej generácie: ako AI mení pracovný priestor.....	451
Otvorené údaje a uberizácia sú hrozboou pre existujúce stavebné podniky .....	453
Nevyriešené problémy uberizácie ako posledná šanca využiť čas na transformáciu .....	456
<b>KAPITOLA 11.2. PRAKTICKÁ PRÍRUČKA NA ZAVEDENIE PRÍSTUPU ZALOŽENÉHO NA ÚDAJOCH .....</b>	<b>461</b>
Od teórie k praxi: plán digitálnej transformácie v stavebníctve .....	461
Položenie digitálneho základu: 1-5 krokov k digitálnej zrelosti .....	463
Uvoľnenie potenciálu údajov: 5-10 krokov k digitálnej zrelosti.....	467
Plán transformácie: od chaosu k spoločnosti riadenej údajmi .....	474
Budovanie v priemysle 5.0: Ako zarábať peniaze, keď sa už nemôžete skrývať .....	477
<b>ZÁVER .....</b>	<b>479</b>
<b>O AUTOROVI .....</b>	<b>482</b>
<b>ĎALŠÍ VZŤAH .....</b>	<b>483</b>
<b>KOMENTÁR K PREKLADU .....</b>	<b>483</b>
<b>ĎALŠIE ZRUČNOSTI A POJMY .....</b>	<b>484</b>
<b>SLOVNÍČEK .....</b>	<b>488</b>
<b>ZOZNAM ODKAZOV A ONLINE MATERIÁLOV .....</b>	<b>495</b>
<b>INDEX PREDMETOV .....</b>	<b>512</b>



## MAXIMÁLNE POHODLIE S TLAČENOU VERZIOU

Držíte v rukách bezplatnú digitálnu verziu knihy **Data-Driven Construction**. Pre pohodlnejšiu prácu a rýchly prístup k materiálom vám odporúčame venovať pozornosť **tlačenému vydaniu**:



■ **Vždy po ruke:** kniha v tlačenej podobe sa stane spoľahlivým pracovným nástrojom, ktorý vám umožní rýchlo nájsť a použiť správne vizualizácie a diagramy v akejkoľvek pracovnej situácii

■ **Vysoká kvalita ilustrácií:** všetky obrázky a grafika v tlačenom vydani sú prezentované v maximálnej kvalite

■ **Rýchly prístup k informáciám:** jednoduchá navigácia, možnosť robiť si poznámky, záložky a pracovať s knihou kdekoľvek.

Zakúpením plnej tlačenej verzie knihy získate praktický nástroj na pohodlnú a efektívnu prácu s informáciami: možnosť rýchlo využívať vizuálne materiály pri každodenných úlohách, rýchlo nájsť potrebné schémy a robiť si poznámky. Okrem toho svojou kúpou podporíte šírenie otvorených poznatkov.

Tlačenú verziu knihy si môžete objednať na adrese: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## I ČASŤ

# **OD HLINENÝCH TABULIEK PO DIGITÁLNU REVOLÚCIU: AKO SA VYVÍJALI INFORMÁCIE V STAVEBNÍCTVE**

V prvej časti knihy sa skúma historický vývoj správy údajov v stavebníctve od primitívnych záznamov na fyzických nosičoch až po moderné digitálne ekosystémy. Analyzuje transformáciu technológií správy informácií, vznik ERP - systémov a vplyv fragmentácie údajov na efektívnosť podnikových procesov. Osobitná pozornosť sa venuje procesu digitalizácie informácií a rastúcemu významu objektívnej analýzy namiesto subjektívneho odborného posúdenia. Podrobne sa skúma exponenciálny nárast objemu informácií, ktorému čelí moderné stavebníctvo, a s tým súvisiace výzvy pre podnikové systémy. Skúma sa postavenie stavebníctva v kontexte štvrtej a piatej priemyselnej revolúcie, ako aj potenciál využitia umelej inteligencie a dátovo orientovaných prístupov na vytvorenie udržateľnej konkurenčnej výhody.

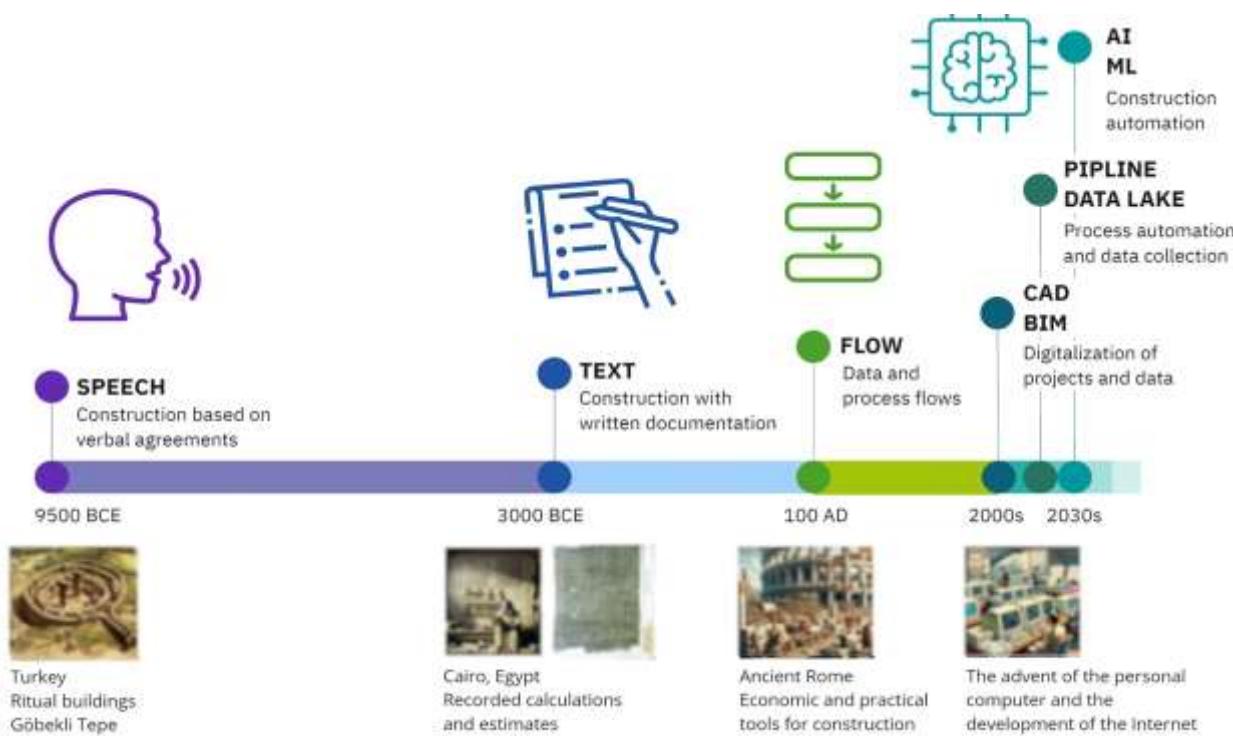
## KAPITOLA 1.1.

# VÝVOJ POUŽÍVANIA ÚDAJOV V STAVEBNÍCTVE

### Zrod dátovej éry v stavebníctve

Približne pred 10 000 rokmi, v období neolitu, ľudstvo uskutočnilo prevratnú zmenu vo svojom vývoji, keď opustilo kočovný spôsob života v prospech usadlého, čo viedlo k vzniku prvých primitívnych stavieb z hliny, dreva a kameňa [6]. Od tohto momentu sa začínajú dejiny stavebníctva.

S rozvojom civilizácií sa architektúra stávala čoraz zložitejšou a vznikali prvé rituálne chrámy a verejné budovy. Zvyšujúca sa zložitosť architektonických návrhov si vyžadovala, aby inžinieri a manažéri v staroveku vytvorili prvé záznamy a výpočty. Prvé záznamy na hlinených tabuľkách a papyrusoch často obsahovali opis logiky výpočtu množstva potrebných stavebných materiálov, ich ceny a výpočtu platby za vykonanú prácu [7]. Tak sa začala éra využívania údajov v stavebníctve - dlho pred nástupom moderných digitálnych technológií (obrázok 1.1-1).

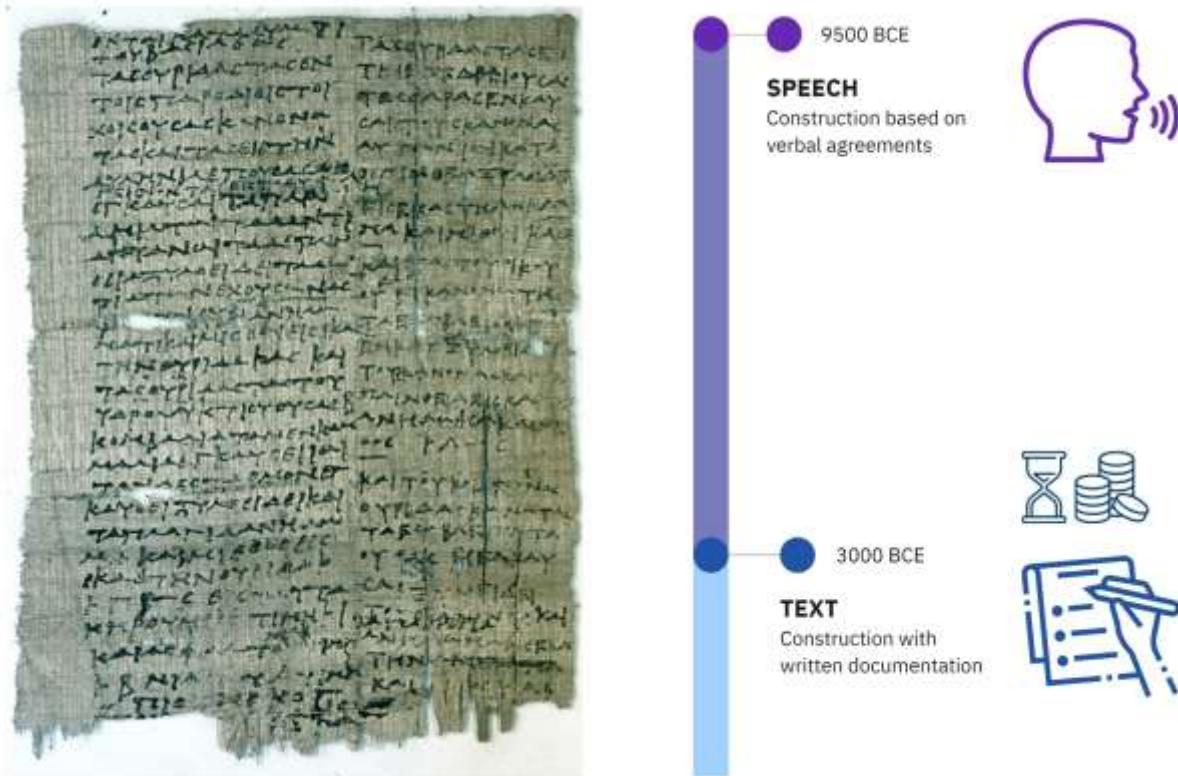


Obr. 1.1-1 Chronológia vývoja informačných technológií v stavebníctve: od verbálnych informácií k umelej inteligencii.

## Od hliny a papyrusu k digitálnej technológií

Prvé doklady o stavbe pochádzajú z obdobia stavby pyramíd, približne 3000-4000 rokov pred našim letopočtom[7]. Odvtedy vedenie písomných záznamov uľahčovalo a sprevádzalo pokrok v stavebníctve a umožnilo zhromažďovanie a systematizáciu poznatkov, ktoré v priebehu nasledujúcich 10 000 rokov viedli k významným inováciám v stavebných metódach a architektúre.

Použitie prvých fyzických médií v stavebníctve, ako sú hlinené tabuľky, papyrus spred tisícročí (obrázok 1.1-2) alebo papier "A0" v 80. rokoch 20. storočia, na zaznamenávanie údajov pôvodne nebolo určené na použitie týchto informácií v nových projektoch. Hlavným účelom takýchto záznamov bolo podrobne informovať o aktuálnom stave projektu vrátane výpočtov potrebných materiálov a nákladov na práce. Podobne aj v dnešnom svete dostupnosť digitálnych projektových údajov a modelov nie vždy zaručuje ich uplatnenie v budúcich projektoch a často slúžia najmä ako informácie pre aktuálne výpočty potrebných materiálov a nákladov na výstavbu.



Obr. 1.1-2 Papyrus z 3. storočia pred n. l. opisujúci náklady na maľovanie rôznych typov okien v kráľovskom paláci technikou enkaustiky.

Ľudstvu trvalo približne 5 000 rokov, kým prešlo od ústnych rozhovorov k písomným dokumentom v oblasti riadenia stavebných projektov, a rovnaký čas potrebovalo na prechod od papierových k digitálnym údajom ako hlavnému zdroju plánovania a kontroly.

Tak ako rozvoj obchodných a peňažných vzťahov podnietil vznik písma a prvých právnikov na riešenie sporov, tak aj prvé záznamy o materiálových nákladoch a rozsahu prác v stavebníctve viedli k vzniku prvých manažérov v stavebníctve, ktorých povinnosťou bolo dokumentovať, monitorovať a zodpovedať za kľúčové informácie o časovom harmonograme projektu a nákladoch.

V súčasnosti zohrávajú údaje oveľa významnejšiu úlohu: nielenže zaznamenávajú prijaté rozhodnutia, ale stávajú sa aj nástrojom na predpovedanie a modelovanie budúcnosti. To je základ, na ktorom je postavený moderný procesný prístup v projektovom riadení - premeniť nahromadené skúsenosti na systém rozhodovania založený na štruktúrovaných a overiteľných údajoch.

## Proces ako nástroj pre skúsenosti založené na údajoch

Základom každého procesu je transformácia minulých skúseností na nástroj plánovania budúcnosti. Skúsenosti v modernom zmysle slova predstavujú štruktúrovaný súbor údajov, ktoré možno analyzovať a na základe ktorých možno robiť kvalifikované predpovede.

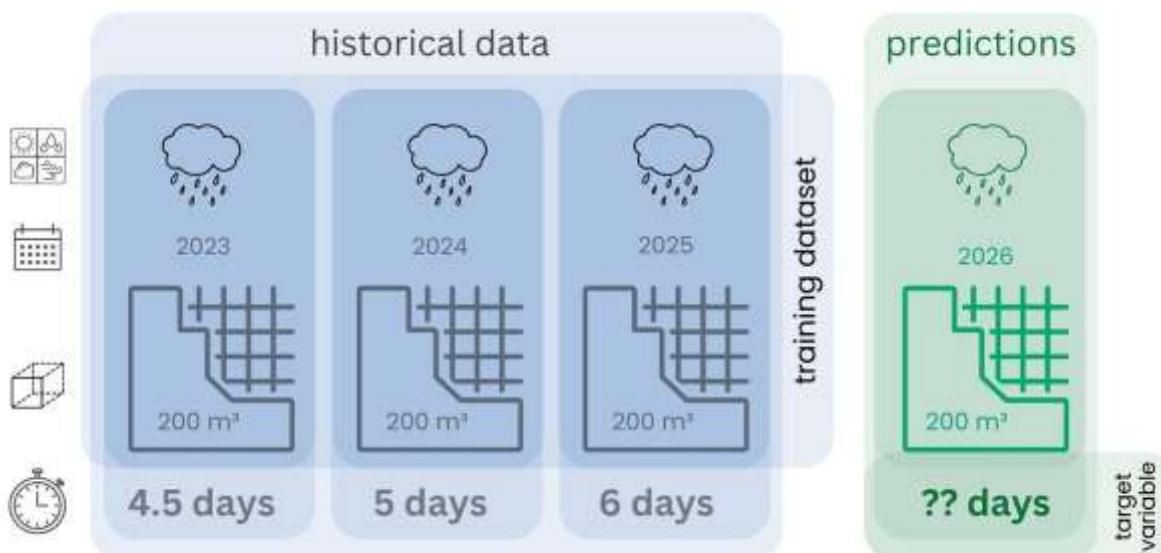
Práve historické údaje slúžia ako základ prognózovania, pretože jasne ukazujú výsledky vykonanej práce a poskytujú prehľad o faktoroch, ktoré tieto výsledky ovplyvňujú.

Vezmime si konkrétny príklad z monolitickej výstavby: pri plánovaní harmonogramu prác sa zvyčajne berie do úvahy objem betónu, zložitosť konštrukcie a poveternostné podmienky. Predpokladajme, že konkrétny stavbyvedúci alebo historické údaje spoločnosti za posledné tri roky (2023 - 2025) ukazujú, že vyliatie monolitickej konštrukcie s plochou 200 m<sup>2</sup> v daždivom počasí trvalo 4,5 až 6 dní (obr. 1.1-3). Práve tieto nahromadené štatistické údaje sa stávajú základom pre predpovedanie času realizácie a kalkuláciu nákladov na zdroje pri plánovaní podobných prác v budúcich projektoch. Na základe týchto historických údajov môže majster alebo odhadca na základe skúseností urobiť informovanú predpoveď času potrebného na dokončenie budúcich podobných prác v roku 2026 za podobných podmienok.

V tomto prípade odhadu času funguje analytický proces ako mechanizmus na transformáciu rôznorodých údajov do štruktúrovaných skúseností a následne do presného plánovacieho nástroja. Údaje a procesy predstavujú jednotný ekosystém, kde jedno bez druhého nemôže existovať.

*Spočítajte, čo sa dá spočítať, zmerajte, čo sa dá zmerať, a urobte merateľným to, čo sa nedá zmerať [8].*

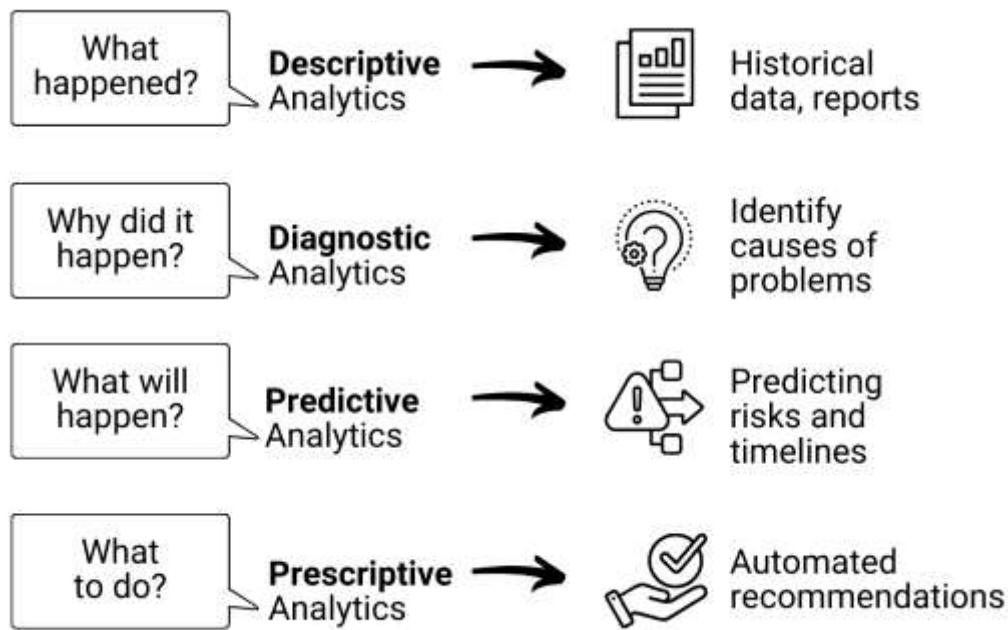
- Galileo Galilei



Obr. 1.1-3 Historické údaje slúžia ako tréningový súbor údajov na predpovedanie jednej z hodnôt v budúcnosti

V dnešnom podnikateľskom prostredí sa analýza údajov stáva rozhodujúcou súčasťou efektívneho riadenia projektov, optimalizácie procesov a strategického rozhodovania. Stavebníctvo si postupne osvojuje štyri kľúčové úrovne analytiky, z ktorých každá odpovedá na špecifickú otázku a poskytuje jedinečné výhody (obr. 1.1-4):

- **Deskriptívna analýza** - odpovedá na otázku "čo sa stalo?" a poskytuje historické údaje a správy o minulých udalostiach a výsledkoch: za posledné tri roky (2023-2025) trvalo vyliatie monolitickej konštrukcie s plochou  $200 \text{ m}^2$  za daždivého počasia 4,5 až 6 dní.
- **Diagnostická analýza** - odpovedá na otázku "prečo sa to stalo?" tým, že identifikuje príčiny problémov: z analýzy vyplýva, že čas betonáže monolitickej konštrukcie sa predĺžil v dôsledku daždivého počasia, ktoré spomalilo proces tvrdnutia betónu
- **Prediktívna analýza** - orientovaná na budúnosť, predpovedanie možných rizík a času realizácie prostredníctvom odpovede na otázku "čo sa stane?": na základe historických údajov sa predpokladá, že vyliatie podobnej monolitickej konštrukcie s rozlohou  $200 \text{ m}^2$  v daždivom počasí v roku 2026 bude trvať približne 5,5 dňa, pričom sa zohľadnia všetky známe faktory a trendy.
- **Preskriptívna analýza** - poskytuje automatizované odporúčania a odpovede na otázku "čo robiť?", čím umožňuje spoločnostiam vybrať optimálne kroky: Na optimalizáciu prác sa napríklad odporúča: použiť špeciálne prísady na urýchlenie tvrdnutia betónu v podmienkach vysokej vlhkosti; plánovať betonáž na obdobie s najnižšou pravdepodobnosťou zrážok; zabezpečiť dočasné prístrešky pre konštrukciu, ktoré skrátia čas prác na 4 až 4,5 dňa aj v nepriaznivých poveternostných podmienkach.



Obr. 1.1-4 Hlavné typy analytiky: od popisu minulosti po automatizované rozhodovanie.

Plnohodnotná digitálna transformácia, ktorá znamená prechod na systémovú analytiku a riadenie založené na údajoch, si vyžaduje nielen outsourcing, ale aj vytvorenie kompetentného interného tímu. Kľúčovými členmi takého tímu by mali byť produktoví manažéri, dátoví inžinieri, analytici a vývojári, ktorí budú úzko spolupracovať s obchodnými jednotkami (obr. 4.3-9). Táto spolupráca je nevyhnutná na kladenie inteligentných analytických otázok a efektívne parametrizovanie úloh obchodného rozhodovania. V informačnej spoločnosti sa údaje nestanú len pomocným nástrojom, ale základom prognózovania a optimalizácie.

Digitálna transformácia v stavebníctve zásadne mení spôsob navrhovania, správy a prevádzky zariadení. Tento proces sa označuje ako digitalizácia informácií - pri ktorej sa všetky aspekty stavebného procesu digitalizujú do digitálnej podoby vhodnej na analýzu.

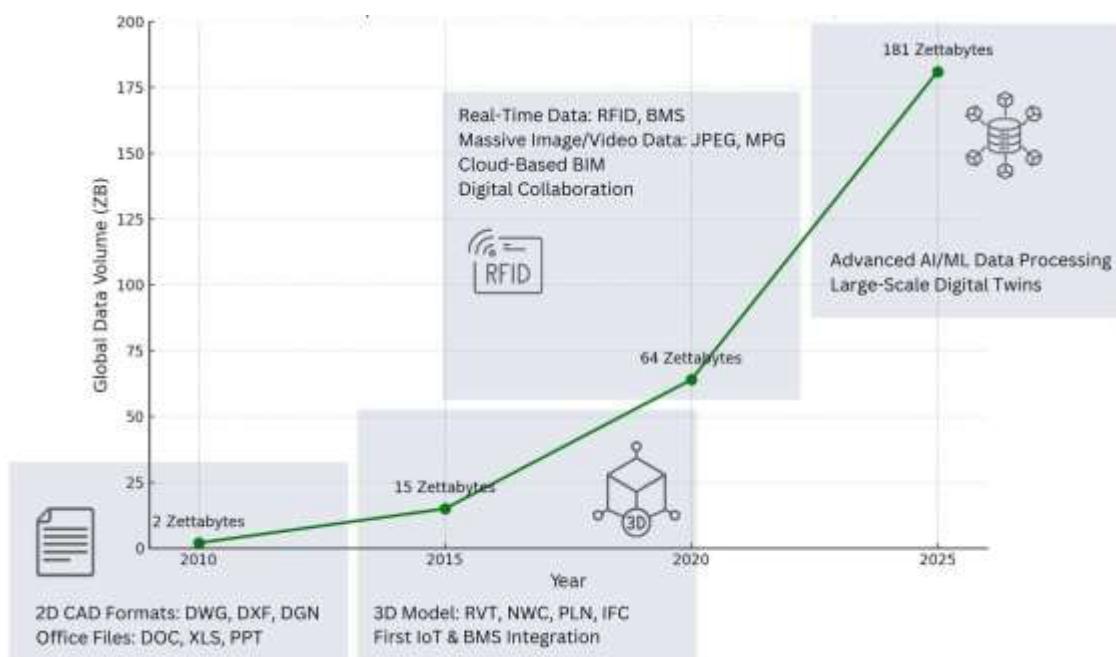
## Digitalizácia informácií o stavebnom procese

Po tisícročia sa množstvo informácií zaznamenaných v stavebníctve takmer nemenilo, ale v posledných desaťročiach rýchlo narastlo (obrázok 1.1-5).

Podľa štúdie spoločnosti PwC® "Managed Data. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] bolo 90% všetkých údajov na svete vytvorených v posledných dvoch rokoch (od roku 2015). Väčšina spoločností však tieto údaje nevyužíva v plnej miere, pretože buď zostávajú v izolovaných systémoch, alebo sa jednoducho archivujú bez skutočnej analýzy.

Nárast objemu dát sa v posledných rokoch len zrýchnil a zdvojnásobil sa z 15 zettabajtov v roku 2015

na 181 zettabajtov v roku 2025 [10]. Denne sa servery stavebných a projektových spoločností zapĺňajú projektovou dokumentáciou, harmonogramami prác, výpočtami a kalkuláciami, finančnými správami. Na 2D/3D -výkresy sa používajú formáty DWG, DXF a DGN a na 3D modely formáty RVT, NWC, PLN a IFC™. Textové dokumenty, tabuľky a prezentácie sa ukladajú vo formátoch DOC, XLSX a PPT. Okrem videa a obrázkov zo staveniska - vo formátoch MPG a JPEG, sa v reálnom čase ukladajú údaje z komponentov internetu vecí, tagov RFID® (identifikácia a sledovanie) a systémov riadenia budov BMS (monitorovanie a riadenie)



Obr. 1.1-5 Parabolický rast údajov 2010-2025 (na základe [10]).

S rýchlym rastom informácií čeli stavebníctvo nielen potrebe zhromažďovať a uchovávať údaje, ale aj zabezpečiť ich overovanie, validáciu, merateľnosť a analytické spracovanie. V súčasnosti odvetvie prechádza aktívnu fázou digitalizácie informácií - systematickou transformáciou všetkých aspektov stavebnej činnosti do digitálnej podoby vhodnej na analýzu, interpretáciu a automatizáciu.

Digitalizácia informácií znamená prevziať informácie o všetkých subjektoch a prvkoch stavebného projektu a samotného stavebného procesu - vrátane tých, ktoré sme predtým vôbec nepovažovali za informácie - a previesť ich do dátového formátu, aby sa informácie dali kvantifikovať a ľahko analyzovať.

V kontexte stavebníctva to znamená zachytávať a digitalizovať informácie všetkých prvkoch projektov a všetkých procesoch - od pohybu strojov a ľudí na stavenisku až po poveternostné a klimatické podmienky na stavenisku, aktuálne ceny materiálov a úrokové sadzby centrálnej banky - s cieľom vytvárať analytické modely.

*Ak dokážete zmerať to, o čom hovoríte, a vyjadriť to v číslach, potom o danej téme niečo viete. Ak to však nedokážete vyjadriť kvantitatívne, vaše vedomosti sú veľmi obmedzené a neuspokojivé. Môže to byť východiskový bod, ale nie je to úroveň skutočného vedeckého poznania. [11].*

- W. Thomson (lord Kelvin), 1824 - 1907, britský vedec

Digitalizácia informácií ďaleko presahuje tradičný prístup k zhromažďovaniu informácií, pri ktorom sa zaznamenávali len základné ukazovatele, ako sú človekohodiny alebo skutočné náklady na materiál. Dnes možno prakticky akúkoľvek udalosť transformovať na dátový tok vhodný na hĺbkovú analýzu pomocou pokročilých analytických nástrojov a techník strojového učenia. Stavebníctvo prešlo zásadným posunom od papierových výkresov, excelovských tabuľiek a ústnych pokynov k digitálnym systémom (obrázok 1.2-4), v ktorých sa každý prvok projektu stáva zdrojom údajov. Dokonca aj zamestnanci - od inžinierov až po stavebných robotníkov na stavbe - sú teraz vnímaní ako súbor digitálnych premenných a súborov údajov.

Podľa štúdie spoločnosti KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023" sa digitálne dvojčatá, umelá inteligencia (AI) a veľké objemy dát, stávajú kľúčovými faktormi zlepšenia ziskovosti projektov [2].

Moderné technológie nielenž zjednodušujú zber informácií a do značnej miery ho automatizujú, ale aj radikálne znižujú náklady na uchovávanie údajov. V dôsledku toho spoločnosti upúšťajú od selektívneho prístupu a uprednostňujú ukladanie celého súboru informácií na neskoršiu analýzu (obr. 2.1-5), čo otvára potenciálne možnosti optimalizácie procesov v budúcnosti.

Digitalizácia informácií a digitalizácia umožňujú odhaliť skryté, doteraz nevyužitú hodnotu informácií. Pri správnej organizácii možno údaje opäťovne použiť, interpretovať a integrovať do nových služieb a riešení.

V budúcnosti digitalizácia informácií pravdepodobne povedie k úplnej automatizácii správy dokumentov, zavedeniu samosprávnych procesov v stavebníctve a vzniku nových profesí - dátových analytikov v stavebníctve, odborníkov na riadenie projektov s umelou inteligenciou a digitálnych inžinierov. Stavebné projekty sa stanú dynamickými zdrojmi informácií a rozhodovanie nebude založené na intuícii alebo subjektívnej skúsenosti, ale na spoľahlivých a reprodukovaných digitálnych faktoch

*Informácie sú ropou 21. storočia a analytika je spaľovací motor [12].*

- Peter Sondergaard, senior viceprezident spoločnosti Gartner®

Podľa štúdie IoT Analytics 2024 [13] sa očakáva, že globálne výdavky na správu a analýzu údajov výrazne vzrastú zo 185,5 miliardy USD v roku 2023 na 513,3 miliardy USD do roku 2030, čo predstavuje 16% ročný nárast. Nie všetky komponenty však rastú rovnakým tempom: rýchlosť rastie analytika, zatiaľ čo rast úložísk sa spomaľuje. Analytika zabezpečí najrýchlejší rast v ekosystéme správy údajov: predpokladá sa, že vzrastie z 60,6 miliardy USD v roku 2023 na 227,9 miliardy USD do roku 2030, čo predstavuje zloženú ročnú mieru rastu 27%.

So zrýchľujúcou sa digitalizáciou informácií a rýchlym rastom ich objemu čelí manažment stavebných projektov a spoločnosti potrebe systematicky uchovávať, analyzovať a spracovávať rôznorodé, často heterogénne údaje. V reakcii na túto výzvu sa od polovice 90. rokov minulého storočia v odvetví začal masívny prechod na elektronickú tvorbu, ukladanie a správu dokumentácie - od tabuľiek a projektových výpočtov až po výkresy a zmluvy.

Tradičné papierové dokumenty, ktoré si vyžadujú podpisy, fyzické uchovávanie, pravidelné revízie a archiváciu v skrinkách, sa postupne nahradzajú digitálnymi systémami, ktoré uchovávajú údaje štruktúrovaným spôsobom - v databázach špecializovaných aplikácií.



## KAPITOLA 1.2.

# TECHNOLÓGIE A SYSTÉMY RIADENIA V MODERНОM STAVEBNÍCTVE

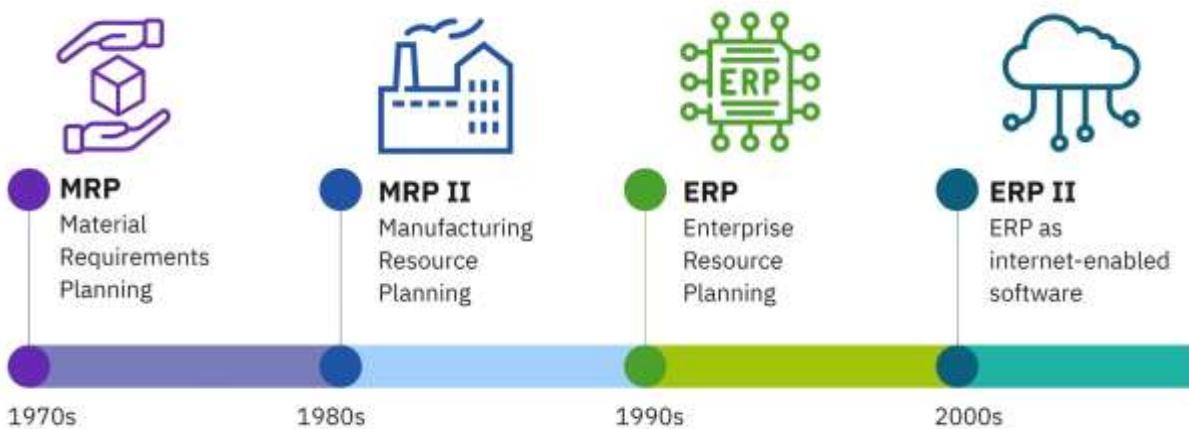
### Digitálna revolúcia a vznik modulárnych MRP/ERP -systémov

Éra moderného digitálneho ukladania a spracovania dát sa začala s príchodom magnetickej pásky v 50. rokoch 20. storočia, ktorá otvorila možnosť ukladania a využívania veľkého množstva informácií. Ďalším prelomom bol príchod diskových jednotiek, ktoré radikálne zmenili prístup k správe dát v stavebníctve.

S rozvojom dátových skladov vstúpilo na trh riešení veľké množstvo spoločností, ktoré začali vyvíjať modulárny softvér na vytváranie, ukladanie, spracovanie údajov a automatizáciu rutinných úloh

Exponenciálny nárast informácií a nástrojov viedol k potrebe integrovaných, modulárnych riešení, ktoré nepracujú s jednotlivými súbormi, ale pomáhajú spravovať a kontrolovať tok údajov v rámci procesov a projektov.

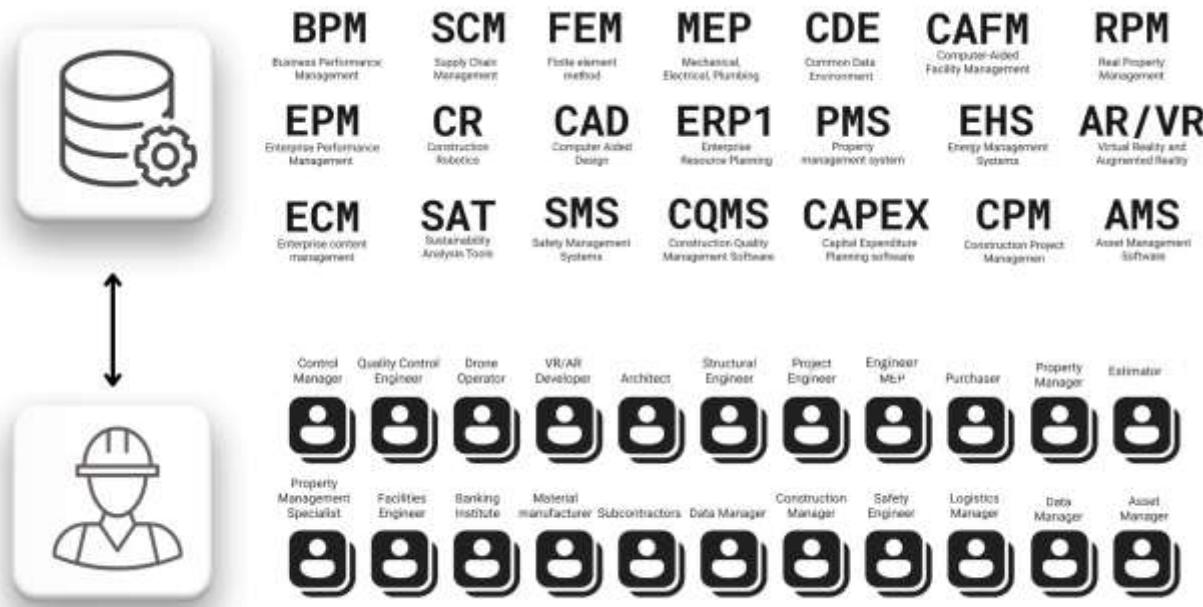
Prvé komplexné nástroje platformy museli nielen ukladať dokumenty, ale aj dokumentovať všetky požiadavky na zmeny a operácie v procesoch: kto ich inicioval, aký bol rozsah požiadavky a čo bolo nakoniec zaznamenané ako hodnota alebo atribút. Na tieto účely bol potrebný systém, ktorý by dokázal sledovať presné výpočty a priaté rozhodnutia (obr. 1.2-1). Takýmito platformami boli prvé systémy MRP (Material Requirements Planning) a ERP (Enterprise Resource Planning), ktoré sa stali populárnymi od začiatku 90. rokov [14]



Obrázok 1.2-1 Pokrok v technológií ukladania údajov viedol v 80. rokoch 20. storočia k vzniku ERP -systémov.

Prvé systémy MRP - a ERP - položili základ éry digitalizácie v oblasti riadenia obchodných procesov a stavebných projektov. Modulárne systémy, pôvodne určené na automatizáciu kľúčových podnikových procesov, boli nakoniec integrované s ďalšími, flexibilnejšími a adaptívnejšími softvérovými riešeniami.

Tieto doplnkové riešenia boli určené na spracovanie údajov a správu obsahu projektu (obr. 1.2-2), budť nahradili niektoré moduly veľkých systémov, alebo ich účinne doplnili, čím rozšírili funkčnosť celého systému.



Obr. 1.2-2 Nové softvérové riešenia prilákali do podniku armádu manažérov, ktorí riadia dátové toky.

V posledných desaťročiach spoločnosti vo veľkom investovali do modulárnych systémov [15] a vnímali ich ako dlhodobé integrované riešenia.

Podľa správy Software Path na rok 2022 [16] je priemerný rozpočet na používateľa systému ERP 9 000 USD. V priemere používa takéto systémy približne 26% zamestnancov spoločnosti. V prípade organizácie so 100 používateľmi tak celkové náklady na implementáciu ERP dosahujú približne 900 000 USD.

Investície do proprietárnych, uzavretých a modulárnych riešení sú čoraz menej opodstatnené na pozadí rýchleho rozvoja moderných, flexibilných a otvorených technológií. Ak sa už takéto investície uskutočnili, je dôležité objektívne prehodnotiť úlohu existujúcich systémov: či zostávajú z dlhodobého hľadiska nevyhnutné, alebo sa ich funkcie dajú revidovať a implementovať efektívnejšie a transparentnejšie.

Jedným z hlavných problémov dnešných modulárnych dátových platform je, že centralizujú správu údajov v rámci uzavretých aplikácií. V dôsledku toho sa údaje - hlavné aktívum spoločnosti - stávajú závislými od konkrétnych softvérových riešení, a nie naopak. To obmedzuje opakovane použitie informácií, komplikuje migráciu na a znižuje agilitu podniku v rýchlo sa meniacom digitálnom prostredí.

Ak je pravdepodobné, že význam alebo dôležitosť uzavretej modulárnej architektúry sa v budúcnosti zníži, má zmysel uznať dnes vynaložené náklady ako utopené náklady a zameriť sa na strategický prechod k otvorenejšiemu, škálovateľnejšiemu a adaptívnejšiemu digitálnemu ekosystému.

Proprietárny softvér sa vyznačuje tým, že vývojár má výlučnú kontrolu nad zdrojovým kódom a údajmi používateľa vytvorenými v rámci používania takýchto riešení. Na rozdiel od softvéru s otvoreným zdrojovým kódom nemajú používatelia prístup k vnútornej štruktúre aplikácie a nemôžu ju nezávisle preskúmať, upraviť alebo prispôsobiť svojim potrebám. Namiesto toho sú povinní zakúpiť si licencie, ktoré im poskytujú právo používať softvér v rámci obmedzení stanovených dodávateľom.

Moderný prístup zameraný na údaje ponúka inú paradigmu: údaje by sa mali považovať za hlavné strategické aktívum - nezávislé, trvalé a oddelené od konkrétnych softvérových riešení. Aplikácie sa zasa stávajú len dátovými nástrojmi, ktoré možno voľne nahradíť bez rizika straty kritických informácií.

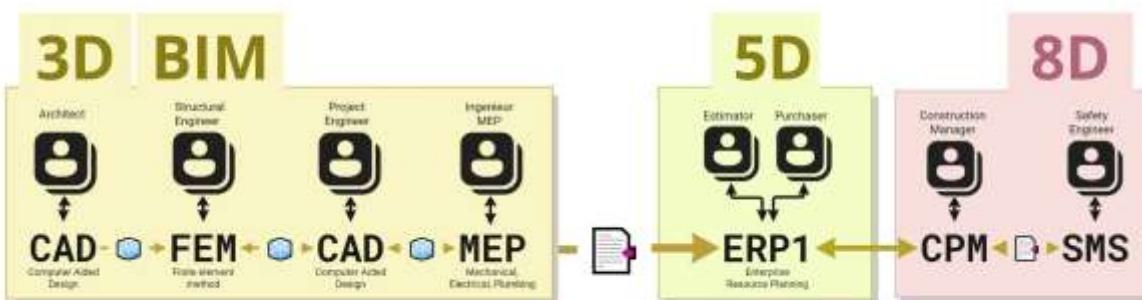
Vývoj systémov ERP a MRP v 90. rokoch 20. storočia (obrázok 1.2-1) poskytol podnikom výkonné nástroje na riadenie procesov, ale mal aj nezamýšľaný dôsledok, ktorým bolo výrazné zvýšenie počtu ľudí zapojených do udržiavania informačných tokov. Namiesto automatizácie a zjednodušenia prevádzkových úloh tieto systémy často vytvárali nové úrovne zložitosti, byrokracie a závislosti od interných IT zdrojov.

## Systémy správy údajov: od dolovania údajov k obchodným výzvam

Dnešné spoločnosti čelia potrebe integrácie viacerých systémov správy údajov. Výber systémov na správu údajov, rozumná správa týchto systémov a integrácia rôznorodých zdrojov údajov sa stávajú rozhodujúcimi pre výkonnosť podniku.

V polovici 20. storočia môžete nájsť stovky (a vo veľkých stavebných spoločnostiach tisíce) rôznych systémov (obrázok 1.2-3), ktoré musia fungovať v súlade, aby všetky aspekty stavebného procesu prebiehali hladko a súdržne.

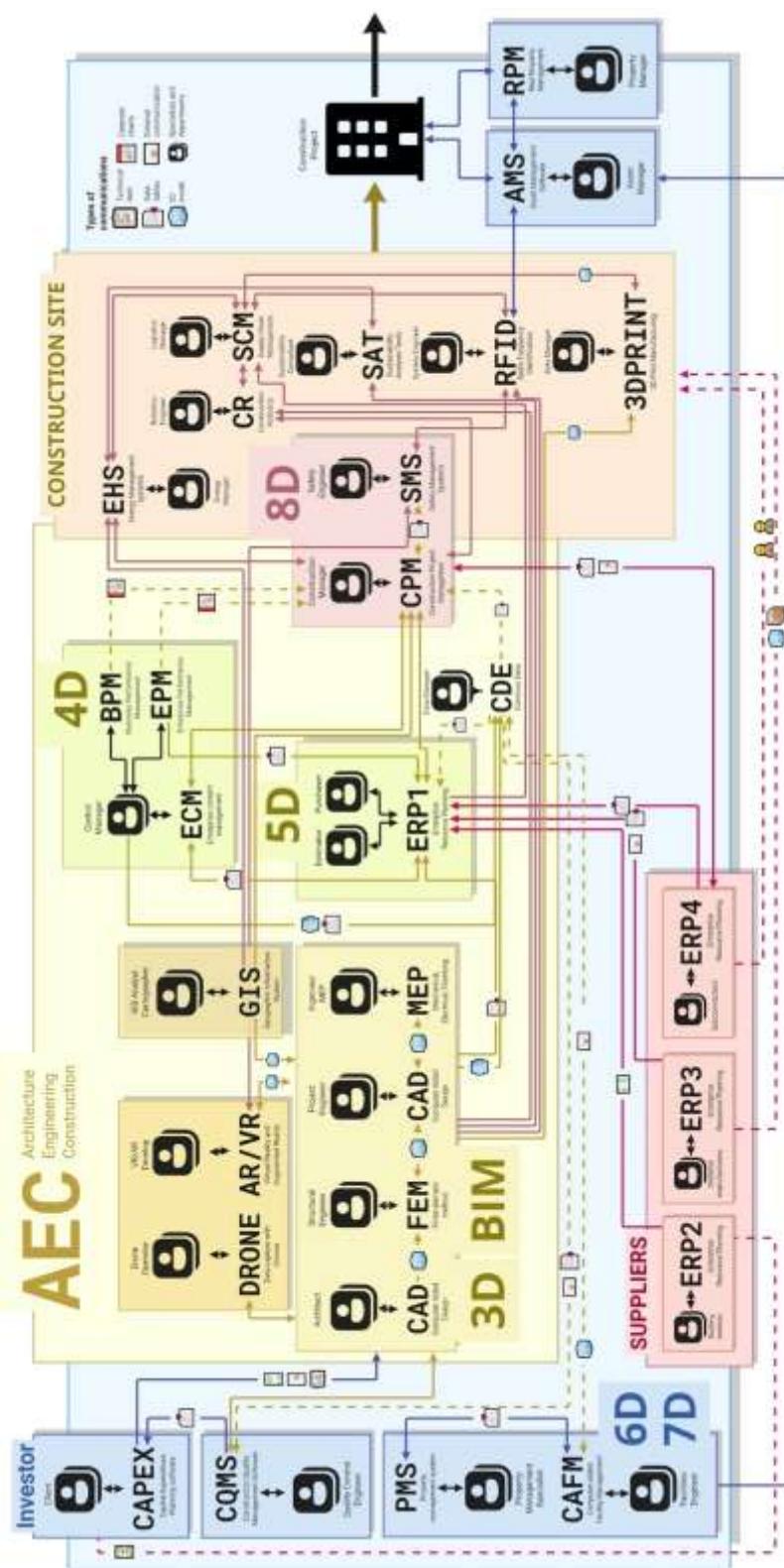
Podľa štúdie spoločnosti Deloitte z roku 2016® "Data-Driven Management in Digital Capital Projects" (Riadenie založené na údajoch v digitálnych kapitálových projektoch) používa priemerný stavebný odborník denne 3,3 softvérových aplikácií, ale len 1,7 z nich je navzájom integrovaných [17].



Obr. 1.2-3 Každý podnikový systém si vyžaduje profesionálny tím a zodpovedného manažéra pre kvalitnú správu údajov.

Nasleduje zoznam populárnych systémov pre stredné a veľké spoločnosti v stavebníctve, ktoré sa používajú pri efektívnom riadení stavebných projektov:

- **ERP (plánovanie podnikových zdrojov)** - poskytuje integráciu obchodných procesov vrátane účtovníctva, obstarávania a riadenia projektov.
- **CAPEX (softvér na plánovanie kapitálových výdavkov)** - slúži na rozpočtovanie a riadenie finančných investícii do stavebných projektov, pomáha určiť náklady na dlhodobý majetok a investície do dlhodobého majetku.
- **CAD (Computer-Aided Design) a BIM (Building Information Modeling)** - sa používajú na vytváranie podrobných a presných technických výkresov a 3D -modelov projektov. Tieto systémy sa zameriavajú na prácu s geometrickými informáciami.
- **MEP (Mechanical, Electrical, Plumbing)** - inžinierske systémy, ktoré zahŕňajú mechanické, elektrické a vodovodné komponenty a podrobne popisujú vnútorný "obehový" systém projektu.
- **GIS (geografické informačné systémy)** - používajú sa na analýzu a plánovanie terénu vrátane kartografie a priestorovej analýzy.
- **CQMS (softvér na riadenie kvality v stavebníctve)** - zabezpečuje súlad stavebných procesov so stanovenými normami a predpismi, čím pomáha eliminovať chyby.
- **CPM (riadenie stavebných projektov)** - zahŕňa plánovanie, koordináciu a kontrolu stavebných procesov.
- **CAFM (Computer-Aided Facility Management)** - systémy správy a údržby budov.
- **SCM (Supply Chain Management)** je potrebný na optimalizáciu toku materiálov a informácií medzi dodávateľmi a staveniskom.
- **EPM (Enterprise Performance Management)** - zameraný na zlepšenie podnikových procesov a výkonnosti.
- **AMS (softvér na správu aktív)** - používa sa na optimalizáciu používania, správy a údržby zariadení a infraštruktúry počas celého životného cyklu aktív.
- **RPM (Správa nehnuteľností)** - zahŕňa úlohy a procesy súvisiace so správou a prevádzkou budov a pozemkov, ako aj súvisiacich zdrojov a majetku.



Obr. 1.2-4 Vzájomné prepojenie systémov, ktoré spája procesy v spoločnosti s tokom informácií medzi jednotlivými oddeleniami.

- **CAE (Computer-Aided Engineering)** - Počítačom podporované inžinierstvo zahŕňa výpočtové a simulačné systémy, ako je analýza konečných prvkov (FEA) a počítačová dynamika tekutín (CFD).
- **CFD (Computational Fluid Dynamics)** - výpočtová dynamika tekutín, modelovanie prúdenia tekutín a plynov. Podkategória CAE.
- **CAPP (počítačom podporované plánovanie procesov)** - počítačom podporované plánovanie procesov. Používa sa na vytváranie máp trás a procesov.
- **CAM (Computer-Aided Manufacturing)** - počítačom podporovaná výroba, tvorba riadiacich programov pre CNC stroje.
- **PDM (Product Data Management)** - Správa údajov o výrobku, systém na ukladanie a správu technickej dokumentácie.
- **MES (Manufacturing Execution System)** je systém riadenia výrobných procesov v reálnom čase.
- **PLM (Product Lifecycle Management)** - riadenie životného cyklu projektového prvku, integruje PDM, CAPP, CAM a ďalšie systémy pre kompletnejšiu kontrolu výrobku od vývoja až po jeho likvidáciu.

Tieto a mnohé ďalšie systémy, ktoré obsahujú rôzne softvérové riešenia, sa stali neoddeliteľnou súčasťou moderného stavebníctva (obr. 1.2-4). Podstatou takýchto systémov sú špecializované databázy s intuitívnymi rozhraniami, ktoré umožňujú efektívne zadávanie, spracovanie a analýzu informácií vo všetkých fázach projektovania a výstavby. Vzájomná integrácia digitálnych nástrojov pomáha nielen optimalizovať pracovné procesy, ale aj výrazne zvyšuje presnosť rozhodnutí, čo má pozitívny vplyv na čas a kvalitu realizácie projektu.

V polovici prípadov však nedochádza k integrácii. Podľa štatistik je len každá druhá aplikácia alebo systém integrovaný s inými riešeniami [17]. To poukazuje na pretrvávajúcu fragmentáciu digitálneho prostredia a zdôrazňuje potrebu vytvoriť otvorené štandardy a jednotné rozhrania na zabezpečenie komplexnej výmeny informácií v rámci stavebného projektu.

Jednou z hlavných výziev v oblasti integrácie pre moderné spoločnosti zostáva vysoká zložitosť digitálnych systémov a požiadavky na kompetencie používateľov potrebné na efektívne vyhľadávanie a interpretáciu informácií. Na podporu každého systému implementovaného v podniku sa vytvára tím špecialistov na čele s kľúčovým manažérom (obr. 1.2-2).

Kľúčový manažér systému zohráva rozhodujúcu úlohu pri správnom smerovaní toku údajov a zodpovedá za kvalitu výsledných informácií, podobne ako boli prví manažéri pred tisícami rokov zodpovední za čísla napísané na papyrus alebo hlinené tabuľky.

Na to, aby sa rôznorodé informačné toky stali nástrojom riadenia, je nevyhnutná schopnosť systematicky integrovať a spravovať údaje. V tejto architektúre musia manažéri vystupovať ako prvky jednotnej siete - ako mycélium, ktoré spája jednotlivé časti spoločnosti do uceleného živého organizmu schopného prispôsobovať sa a vyvíjať.

## **Firemné mycélium: ako sa údaje spájajú s podnikovými procesmi**

Proces integrácie údajov do aplikácií a databáz je založený na agregácii informácií z rôznych zdrojov vrátane rôznych oddelení a špecialistov (obrázok 1.2-4). Špecialisti vyhľadávajú relevantné údaje, spracúvajú ich a prenášajú do svojich systémov a aplikácií na ďalšie použitie.

Každý podnikový systém pozostávajúci zo súboru nástrojov, technológií a databáz je stromom znalostí, ktorý je zakorenéný v pôde historických údajov a rastie, aby priniesol nové ovocie v podobe hotových riešení: dokumentov, výpočtov, tabuľiek, grafov a informačných panelov (obr. 1.2-5). Systémy v podniku, podobne ako stromy v určitom kúsku lesa, navzájom interagujú a komunikujú, pričom tvoria komplexný a dobre štruktúrovaný systém podporovaný a riadený odbornými manažérmi.

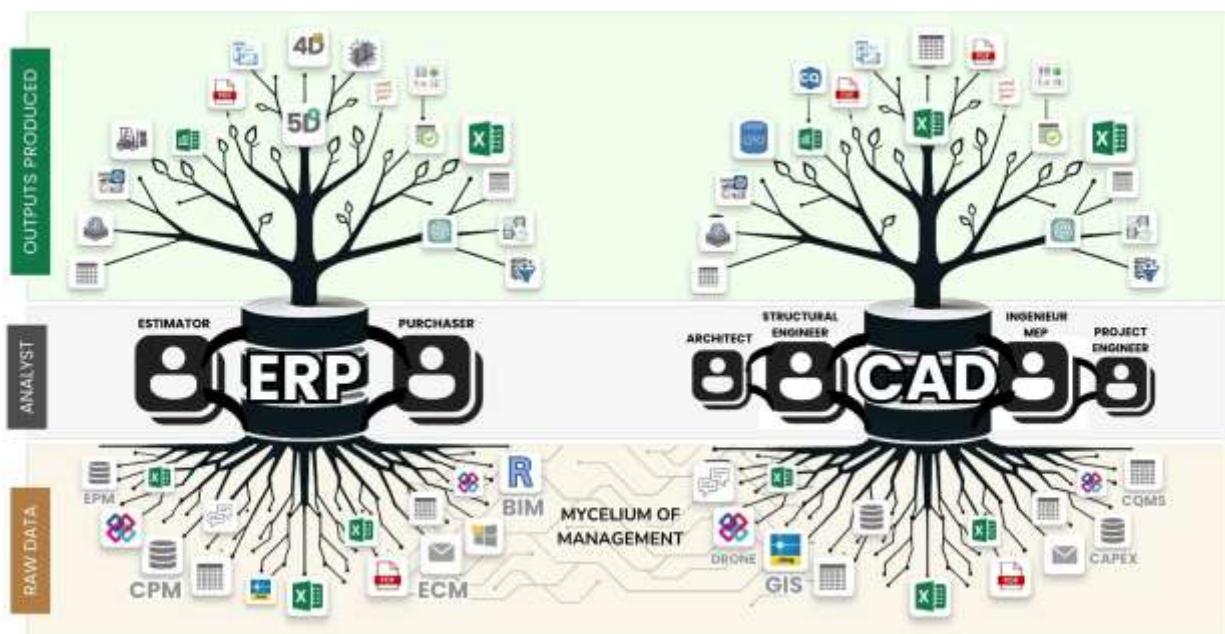
Systém získavania a prenosu informácií v spoločnosti funguje ako komplexná lesná sieť pozostávajúca zo stromov (systémov) a hub mycélia (manažérov), ktoré fungujú ako vodiče a recyklátory a zabezpečujú prenos informácií a ich tok do správnych systémov. To pomáha udržiavať zdravý a efektívny tok a distribúciu údajov v rámci spoločnosti.

Odborníci, podobne ako korene, absorbujú surové údaje v počiatocných fázach projektu a premieňajú ich na živiny pre podnikový ekosystém. Systémy na správu údajov a obsahu (obr. 1.2-4 - ERP, CPM, BIM, atď.) fungujú ako výkonné informačné diaľnice, ktorými tieto znalosti cirkulujú na všetkých úrovniach spoločnosti.

Podobne ako v prírode, kde každý prvok ekosystému zohráva svoju úlohu, aj v podnikovom prostredí každý účastník procesu - od inžiniera po analyтика - prispieva k rastu a plodnosti informačného prostredia. Tieto systémové "dátové stromy" (obr. 1.2-5) nie sú len mechanizmom na zhromažďovanie informácií, ale konkurenčnou výhodou, ktorá zabezpečuje udržateľnosť podniku.

Lesné ekosystémy sú prekvapivo presným odrazom toho, ako sú organizované digitálne podnikové štruktúry. Podobne ako viacúrovňová štruktúra lesa - od podrastu po koruny stromov - aj podnikové riadenie prideľuje úlohy úrovniam zodpovednosti a funkčným oddeleniam.

Hlboké a rozvetvené korene stromov poskytujú odolnosť a prístup k živinám. Podobne pevná organizačná štruktúra a stabilné procesy práce s kvalitnými údajmi podporujú celý informačný ekosystém spoločnosti a prispievajú k jej udržateľnému rastu a rozvoju aj v obdobiah nestability (silného vetra) a kríz na trhu.



Obr. 1.2-5 Integrácia údajov prostredníctvom rôznych systémov je ako mycélium spájajúce manažérov a špecialistov do jednej informačnej siete.

Moderné chápanie rozsahu v podnikaní sa vyvinulo. Hodnotu podniku dnes neurčuje len jeho viditeľná časť - "korunka" v podobe záverečných dokumentov a správ - ale aj hlbka "koreňového systému" kvalitatívne zozbieraných a systematicky spracovaných údajov. Čím viac informácií sa dá zhromaždiť a spracovať, tým vyššia je hodnota podniku. Spoločnosti, ktoré metodicky zhromažďujú "kompost" už spracovaných údajov a dokážu z nich získať užitočné poznatky, získavajú strategickú výhodu

Historické informácie sa stávajú novým druhom kapitálu, ktorý umožňuje rast, optimalizáciu procesov a konkurenčnú výhodu. Vo svete založenom na údajoch nevyhráva ten, kto má viac, ale ten, kto viac vie.

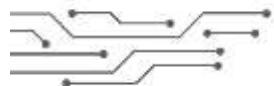
Pre stavebníctvo to znamená prechod na riadenie projektov v reálnom čase, kde všetky procesy - od návrhu a obstarávania až po koordináciu dodávateľov - budú založené na relevantných, denne aktualizovaných údajoch. Integrácia informácií z rôznych zdrojov (ERP-systémy, CAD-modely, senzory IoT na staveniskách, RFID) umožní robiť presnejšie prognózy, rýchlo reagovať na zmeny a predchádzať oneskoreniam spôsobeným nedostatkom aktuálnych údajov.

Podľa publikácie spoločnosti McKinsey & Company Data-Driven Enterprise 2025 (McKinsey & Company®, 2022 [18]) sa úspešné spoločnosti budúcnosti budú spoliehať na údaje vo všetkých kľúčových aspektoch svojej činnosti, od strategických rozhodnutí až po prevádzkové interakcie.

Údaje prestanú byť len analytickým nástrojom a stanú sa neoddeliteľnou súčasťou všetkých obchodných procesov, ktoré zabezpečia transparentnosť, kontrolu a automatizáciu riadenia. Krok založený na údajoch podľa umožní organizáciám minimalizovať vplyv ľudského faktora, znížiť

prevádzkové riziká a zvýšiť transparentnosť a efektívnosť rozhodovania.

21. storočie obracia ekonomickú paradigmu naruby: kým kedysi sa ropa nazývala "čiernym zlatom" pre jej schopnosť poháňať stroje a dopravu, dnes sa historické údaje stlačené pod časovým tlakom stávajú novým strategickým zdrojom, ktorý poháňa nie stroje, ale rozhodovacie algoritmy, ktoré budú riadiť podnikanie.



## KAPITOLA 1.3.

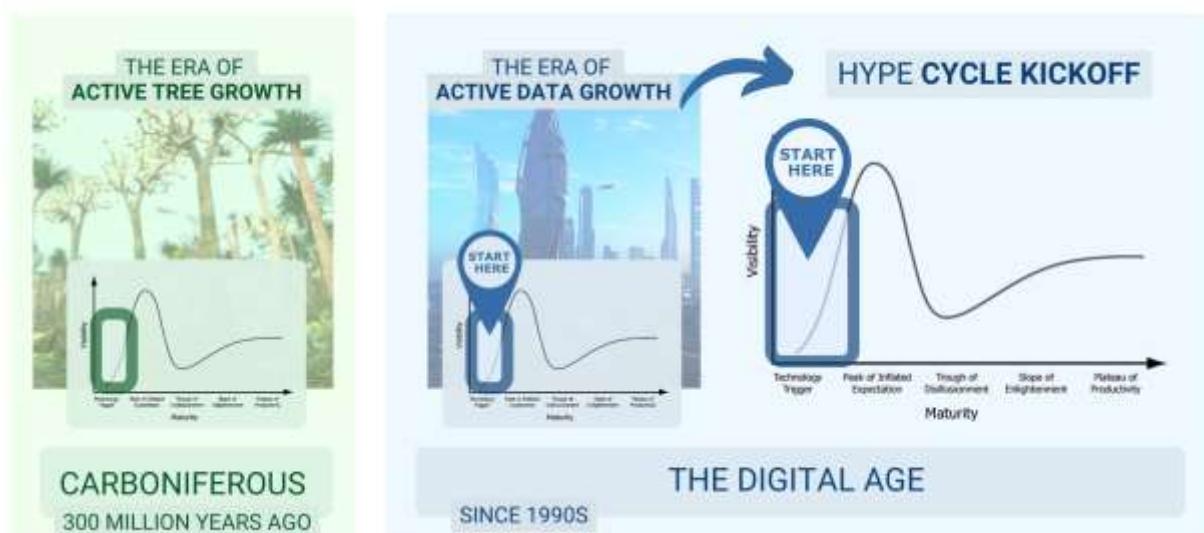
# DIGITÁLNA REVOLÚCIA A EXPLÓZIA DÁT

### Začiatok dátového boomu ako evolučná vlna

Stavebníctvo zažíva bezprecedentnú informačnú explóziu. Ak si predstavíme podnikanie ako strom znalostí (obr. 1.2-5), ktorý je napájaný údajmi, súčasnú fázu digitalizácie možno pirovnať k rýchlemu rastu vegetácie v období karbónu, teda v období, keď sa biosféra Zeme zmenila rýchlym hromadením biomasy (obr. 1.3-1).

S globálnou digitalizáciou sa množstvo informácií v stavebníctve každoročne zdvojnásobuje. Moderné technológie umožňujú zhromažďovať údaje na pozadí, analyzovať ich v reálnom čase a využívať v rozsahu, ktorý sa ešte prednedávnom zdal nemožný.

Podľa Moorovho zákona, ktorý sformuloval Gordon Moore (spoluzakladateľ spoločnosti Intel®), sa hustota a zložitosť integrovaných obvodov a množstvo spracovaných a uložených údajov zdvojnásobuje približne každé dva roky [19].



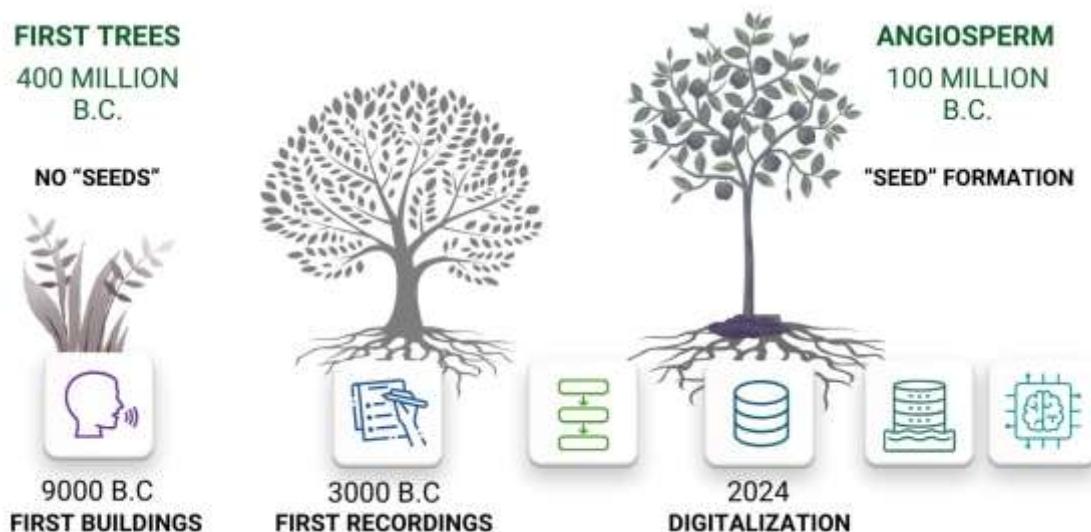
Obr. 1.3-1 Nástup digitalizácie viedol k exponenciálnemu nárastu údajov, podobne ako prudký nárast vegetácie v dobe uhlia.

Zatiaľ čo staroveké megalitické stavby ako Göbekli Tepe (Turecko) nezanechali zdokumentované poznatky vhodné na opäťovné použitie, digitálne technológie dnes umožňujú zhromažďovať a opäťovne používať informácie. Možno to pirovnať k evolučnému prechodu od výtrusných rastlín k semenným rastlinám (rastliny rodu Angiosperm): vznik semen dal podnet k rozsiahlemu rozšíreniu života na planéte. (Obr. 1.3-2).

Podobne sa údaje z minulých projektov stávajú akýmisi "digitálnymi semenami" - nosičmi znalostí DNA, ktoré možno rozšíriť a použiť v nových projektoch a producovať. Nástup moderných nástrojov umelej inteligencie - strojového učenia a veľkých jazykových modelov (LLM), ako sú ChatGPT, LLaMa,

Mistral, Claude, DeepSeek, QWEN, Grok - umožňuje automaticky extrahovať, interpretovať a aplikovať údaje v nových kontextoch

Tak ako semená spôsobili revolúciu v šírení života na pôvodne neživej planéte, "dátové semená" sa stávajú základom pre automatický vznik nových informačných štruktúr a znalostí, čo umožňuje digitálnym ekosystémom vyvíjať sa nezávisle a prispôsobovať sa meniacim sa požiadavkám používateľov.



Obr. 1.3-2 Digitálne "dátové semená" zohrávajú rovnakú evolučnú úlohu ako drevokazné rastliny, kvitnúce rastliny, ktoré zmenili ekosystém Zeme.

Stojíme na prahu novej éry v stavebníctve, kde explózia údajov a aktívne šírenie "dátových semien" - štruktúrovaných informácií z minulých a prebiehajúcich projektov - tvoria základ digitálnej budúcnosti odvetvia. Ich "opel'ovanie" prostredníctvom veľkých dátových jazykových modelov (LLM) nám umožňuje nielen pozorovať digitálne zmeny, ale aktívne sa podieľať na vytváraní samoučiacich sa, adaptívnych ekosystémov. Toto nie je evolúcia - je to digitálna revolúcia, v ktorej sa údaje stávajú hlavným stavebným kameňom novej reality

*Množstvo údajov v stavebníctve dramaticky narastá vďaka informáciám z rôznych odborov počas celého životného cyklu stavebných projektov. Toto obrovské nahromadenie údajov posunulo stavebníctvo do éry veľkých dát [20].*

- Prof. Hang Yang, Katedra stavebného inžinierstva a architektúry, Wuhanská technologická univerzita, Wuhan.,

Rast objemu dát v informačnom veku pripomína evolučné procesy v prírode: tak ako vývoj lesov zmenil dávnu krajinu planéty, súčasná informačná explózia mení krajinu celého stavebníctva.

## Množstvo údajov generovaných v modernej spoločnosti

Za posledné dva roky bolo vytvorených 90% všetkých existujúcich údajov na svete [21]. Od roku 2023 každý človek vrátane odborníkov v stavebnictve vytvorí približne 1,7 megabajtu údajov za sekundu [22] a celkový objem údajov na svete dosiahne v roku 2023 64 zettabajtov a predpokladá sa, že do roku 2025 prekročí 180 zettabajtov, teda  $180 \cdot 10^{15}$  megabajtov [23].

Táto informačná explózia má historický precedens - vynález tlačiarenského stroja Johannesom Gutenbergom v 15. storočí. Len päťdesiat rokov po jeho zavedení sa počet kníh v Európe zdvojnásobil: za niekoľko desaťročí sa vytlačilo toľko kníh, koľko sa vytvorilo ručne za predchádzajúcich 1 200 rokov [24]. Dnes sme svedkami ešte rýchlejšieho rastu: množstvo dát na svete sa zdvojnásobuje každé tri roky.

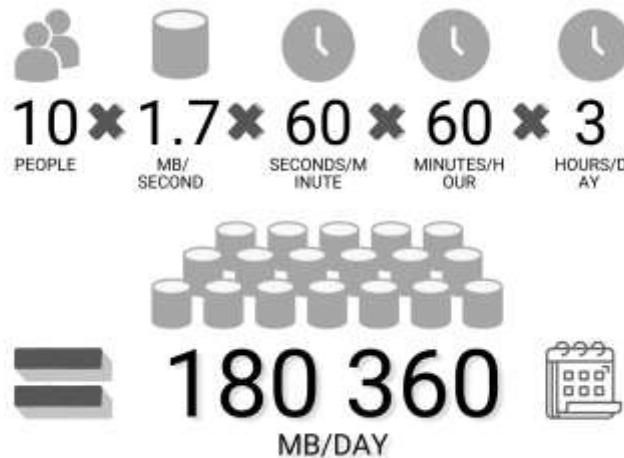
Vzhľadom na súčasné tempo rastu údajov má stavebný priemysel potenciál vygenerovať v nasledujúcich niekoľkých desaťročiach toľko informácií, koľko ich nazbieraných za celú svoju predchádzajúcu história



Obr. 1.3-3 Každodenné ukladanie dát každým zamestnancom na servery spoločnosti prispieva k neustálemu nárastu dát.

V dnešnom svete stavebnictva aj malé spoločnosti denne generujú obrovské množstvo multiformátových informácií a digitálna stopa aj malej stavebnej spoločnosti môže dosiahnuť desiatky gigabajtov denne - od modelov a výkresov až po fotografické záznamy a snímače na stavbe. Ak predpokladáme, že každý technik vygeneruje v priemere približne 1,7 MB údajov za sekundu, zodpovedá to približne 146 GB denne alebo 53 TB ročne (obr. 1.3-3).

Ak tím 10 ľudí aktívne pracuje len 3 hodiny denne, kumulatívne množstvo informácií vytvorených za deň dosiahne 180 gigabajtov (obr. 1.3-4).



Obrázok 1.3-4 Spoločnosť s 10 zamestnancami vygeneruje denne približne 50-200 gigabajtov údajov.

Za predpokladu, že 30% pracovných údajov je nových (zvyšok sa prepíše alebo vymaže), môže firma s 10 zamestnancami vytvoriť rádovo niekoľko stoviek gigabajtov nových údajov mesačne (skutočné čísla závisia od typu činnosti spoločnosti)

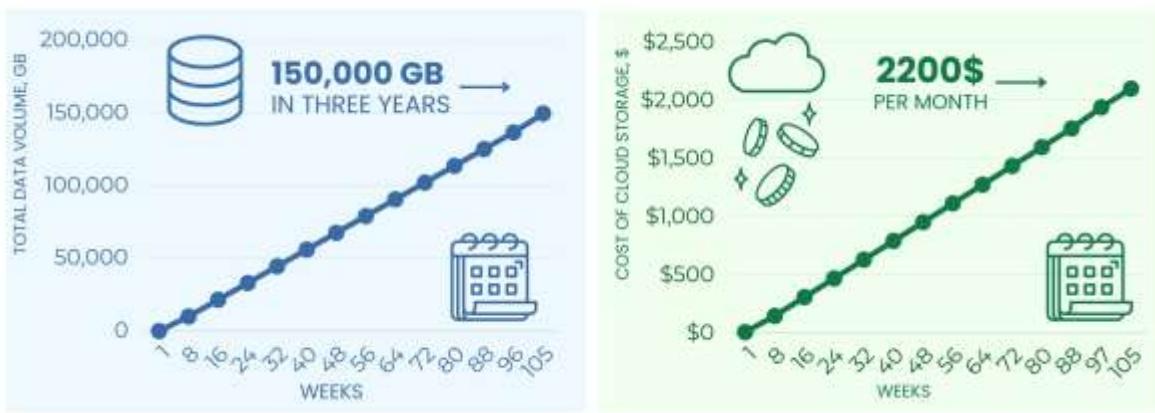
Je teda zrejmé, že nielenže generujeme čoraz viac údajov, ale čelíme aj rastúcej potrebe ich efektívnej správy, ukladania a dlhodobej dostupnosti. A zatiaľ čo predtým mohli údaje "ležať" na lokálnych serveroch bez akýchkoľvek nákladov, v súvislosti s digitálnou transformáciou začína čoraz viac spoločností využívať clouдовé riešenia ako základ svojej informačnej infraštruktúry.

### Náklady na ukladanie údajov: ekonomický aspekt

V posledných rokoch čoraz viac spoločností využíva externé ukladanie údajov prostredníctvom clouдовých služieb. Ak napríklad spoločnosť hostí polovicu svojich údajov v cloude pri priemernej cene 0,015 USD za gigabajt mesačne, jej náklady na ukladanie údajov sa môžu každý mesiac zvýšiť o 10 až 50 USD [25].

Pre malú spoločnosť s typickými vzorcami tvorby údajov sa náklady na clouдовé úložisko môžu pohybovať od stoviek až po potenciálne viac ako tisíc dolárov mesačne (obrázok 1.3-5) v niekoľkých rokoch, čo predstavuje potenciálne značnú finančnú záťaž.

Podľa štúdie spoločnosti Forrester "Enterprises Outsource Data Storage as Complexity Grows" [26], v ktorej sa uskutočnil prieskum medzi 214 vedúcimi pracovníkmi v oblasti technologickej infraštruktúry [26], ktorá sa týkala 214 rozhodovacích pracovníkov v oblasti technologickej infraštruktúry, viac ako tretina organizácií outsourcuje dátové úložisko, aby zvládla rastúci objem a zložitosť dátových operácií, pričom takmer dve tretiny podnikov uprednostňujú model založený na predplatnom.



Obr. 1.3-5 Presun údajov do clodu môže zvýšiť mesačné náklady na ukladanie dát až o 2 000 USD, a to aj v prípade spoločnosti s iba 10 zamestnancami.

Situáciu ďalej komplikuje zrýchlené zavádzanie cloudových technológií, ako sú CAD (BIM), CAFM, PMIS a ERP - systémy, ktoré ďalej zvyšujú náklady na ukladanie a spracovanie údajov. V dôsledku toho sú spoločnosti nútené hľadať spôsoby, ako optimalizovať náklady a znížiť závislosť od poskytovateľov cloudových služieb.

Od roku 2023 sa s aktívnym vývojom veľkých jazykových modelov (LLM) začali meniť prístupy k ukladaniu údajov. Čoraz viac spoločností uvažuje o tom, že by si vzali späť kontrolu nad svojimi údajmi, pretože je bezpečnejšie a výhodnejšie spracovávať informácie na vlastných serveroch.

V tejto súvislosti vystupuje do popredia trend odklonu od cloudového ukladania a spracovania len potrebných údajov v prospech lokálneho nasadenia podnikových riešení LLM a AI. Ako zdôraznil generálny riaditeľ spoločnosti Microsoft v jednom zo svojich rozhovorov [27], namiesto toho, aby sa na vykonávanie rôznych úloh spoliehali na niekoľko samostatných aplikácií alebo cloudových riešení SaaS, budú agenti AI riadiť procesy v databázach a automatizovať funkcie rôznych systémov.

[...] starý prístup k tejto problematike [spracovania údajov] bol takýto: ak si spomeniete, ako rôzne podnikové aplikácie riešili integráciu, používali konektory. Spoločnosti predávali licencie na tieto konektory a na základe toho sa vytvoril obchodný model. SAP [ERP] je jedným z klasických príkladov: k údajom SAP ste mohli pristupovať len vtedy, ak ste mali správny konektor. Takže sa mi zdá, že niečo podobné vznikne aj v prípade interakcie agentov [AI] [...]. Prístup, aspoň ten, ktorý zastávame my, je takýto: myslím si, že koncept existencie obchodných aplikácií sa v ére agentov [AI] pravdepodobne zrúti. Pretože ak sa nad tým zamyslíte, sú to v podstate databázy s množstvom obchodnej logiky

- Satya Nadella, generálny riaditeľ spoločnosti Microsoft, rozhovor pre kanál BG2, 2024. [28]

V tejto paradigme ide prístup LLM založený na údajoch nad rámec klasických systémov. Umelá inteligencia sa stáva sprostredkovateľom medzi používateľom a údajmi (obr. 2.2-3, obr. 2.2-4), čím sa

eliminuje potrebu viacerých sprostredkovateľských rozhraní a zvyšuje sa efektívnosť obchodných procesov. O tomto prístupe k práci s údajmi budeme viac hovoriť v kapitole "Premena chaosu na poriadok a znižovanie zložitosti".

Zatiaľ čo architektúra budúcnosti sa ešte len formuje, spoločnosti už čelia dôsledkom minulých rozhodnutí. Masívna digitalizácia posledných desaťročí sprevádzaná zavádzaním rôznorodých systémov a nekontrolovaným hromadením údajov viedla k novému problému - informačnému preťaženiu.

## Hranice akumulácie údajov: od hmotnosti k významu

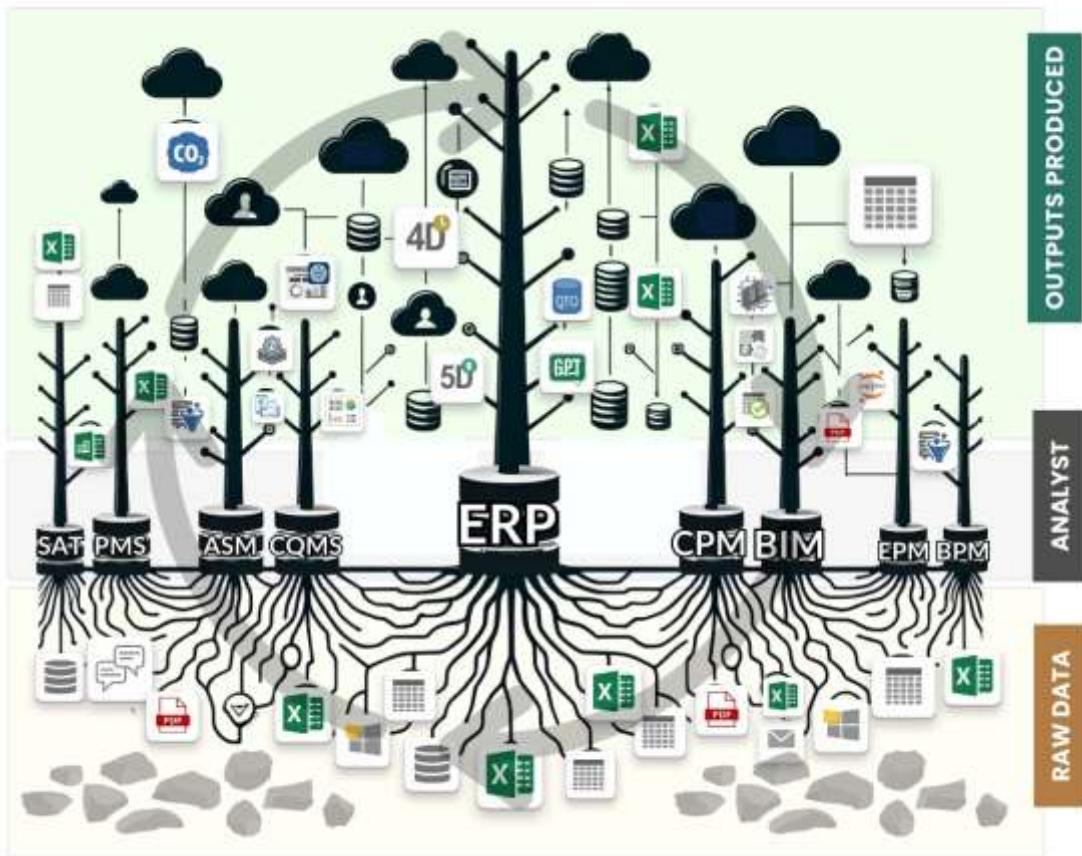
Moderné podnikové systémy sa úspešne rozvíjajú a fungujú v podmienkach riadeného rastu, keď je objem údajov a počet aplikácií v rovnováhe s možnosťami IT oddelení a manažérov. V posledných desaťročiach však digitalizácia viedla k nekontrolovanému nárastu objemu a zložitosti údajov, čo spôsobilo efekt presýtenia informačného ekosystému spoločnosti.

V súčasnosti sú servery a úložiská vystavené bezprecedentnému prílevu nespracovaných a multiformátových informácií, ktoré sa nestihnu premeniť na kompost a rýchlo strácajú význam. Obmedzené podnikové zdroje nedokážu zvládnuť túto záplavu a údaje sa hromadia v izolovaných silách (tzv. "silách"), ktoré si vyžadujú manuálne spracovanie, aby sa z nich získali užitočné informácie.

V dôsledku toho moderné systémy riadenia spoločností často trpia informačným preťažením, podobne ako les zarastený brečtanom a pokrytý plesňou. Namiesto vyživujúceho informačného humusu v srdci podnikového ekosystému sa vytvárajú izolované oblasti údajov rôzneho formátu, čo nevyhnutne vedie k zníženiu celkovej efektívnosti podnikových procesov.

Po dlhom období exponenciálneho rastu dát za posledných 40 rokov bude nevyhnutne nasledovať fáza nasýtenia a následného ochladenia. Keď úložisko dosiahne svoje limity, nastane kvalitatívny posun: údaje už nebudú len objektom úložiska, ale strategickým zdrojom.

S rozvojom umelej inteligencie a strojového učenia majú spoločnosti možnosť znížiť náklady na spracovanie informácií a prejsť od kvantitatívneho rastu ku kvalitatívному využívaniu údajov. V nasledujúcim desaťročí sa bude musieť stavebníctvo presunúť od vytvárania stále väčšieho množstva údajov k zabezpečeniu ich štruktúry, integrity a analytickej hodnoty.



Obrázok 1.3-6 Izolované zdroje údajov bránia zdieľaniu informácií medzi dátovými systémami.

Hlavná hodnota už nespočíva v objeme informácií, ale v schopnosti automaticky ich interpretovať a premeniť na aplikované znalosti užitočné pre manažérské rozhodnutia. Aby sa údaje stali skutočne užitočnými, musia sa správne spravovať: zhromažďovať, overovať, štruktúrovať, uchovávať a analyzovať v kontexte konkrétnych obchodných úloh.

Proces analýzy údajov v spoločnosti sa podobá cyklu života a rozkladu stromov v lese a vzniku nových mladých a silných stromov: zrelé stromy odumierajú, rozkladajú sa a stávajú sa živou pôdou pre nový rast. Ukončené a dokončené procesy sa po dokončení stávajú súčasťou informačného ekosystému spoločnosti a nakoniec sa stanú informačným humusom, ktorý poháňa budúci rast nových systémov a údajov.

V praxi sa však tento cyklus často preruší. Namiesto organickej obnovy vzniká vrstevnatý chaos, podobný geologickým vrstvám, kde sa nové systémy vrstvia na staré bez hlbokej integrácie a štruktúrovania. V dôsledku toho vznikajú rôznorodé informačné "silá", ktoré bránia cirkulácii poznatkov a komplikujú správu údajov.

## Ďalšie kroky: od teórie údajov k praktickej zmene

Vývoj údajov v stavebníctve je cestou od hlinených tabuľiek k moderným modulárnym platformám. Dnešnou výzvou nie je zhromažďovanie informácií, ale vytvorenie rámca, ktorý premení rôznorodé a rozmanité údaje na strategický zdroj. Bez ohľadu na to, či je vašou úlohou manažér spoločnosti alebo inžinier, pochopenie hodnoty údajov a spôsobu práce s nimi bude v budúcnosti klúčovou zručnosťou.

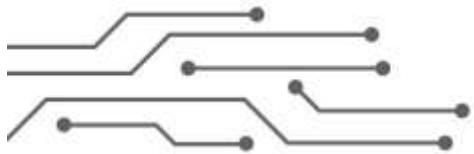
Na zhrnutie tejto časti stojí za to zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované prístupy vo vašich každodenných úlohách:

- Vykonajte osobný audit informačných tokov
  - Vytvorte si zoznam všetkých systémov a aplikácií, s ktorými denne pracujete.
  - Označte, kde trávite najviac času vyhľadávaním alebo dvojitou kontrolou údajov.
  - Identifikujte svoje klúčové zdroje informácií
  - Analýza súčasného prostredia aplikácií z hľadiska nadbytočnosti a duplicity funkcií
- Snažte sa postupovať v procesoch podľa úrovni analytickej vyspelosti
  - Začnite svoje úlohy s popisnou analýzou (čo sa stalo?)
  - Postupne zavádzajte diagnostiku (prečo sa to stalo?)
  - Premýšľajte o tom, ako môžete v procesoch prejsť na prediktívnu (čo sa stane?) a preskriptívnu (čo robiť?) analýzu.
- Začnite štruktúrovať svoje pracovné údaje
  - Zavedenie jednotného systému pomenovania súborov a priečinkov, ktoré často používate pri práci
  - Vytvorenie šablón pre často používané dokumenty a správy
  - Pravidelne archivujte dokončené projekty s jasnou štruktúrou

Aj keď nemôžete zmeniť celú informačnú infraštruktúru vo svojom tíme alebo spoločnosti, začnite s vlastnými procesmi a malými zlepšeniami v každodennej práci. Nezabúdajte, že skutočná hodnota údajov nespočíva v ich objeme, ale v schopnosti získať z nich užitočné poznatky. Aj malé, ale správne štruktúrované a analyzované súbory údajov môžu mať po začlenení do rozhodovacích procesov významný vplyv.

V ďalších častiach knihy prejdeme ku konkrétnym metódam a nástrojom na prácu s údajmi, pozrieme sa na spôsoby transformácie neštruktúrovaných informácií na štruktúrované súbory, preskúmame technológie na automatizáciu analytiky a podrobne popíšeme, ako vybudovať efektívny analytický ekosystém v stavebnej spoločnosti.





## II ČASŤ

# AKO SA STAVEBNÍCTVO TOPÍ V DÁTOVOM CHAOSE

Druhá časť je venovaná kritickej analýze výziev, ktorým čelia stavebné spoločnosti pri práci s rastúcim objemom údajov. Podrobne sa skúmajú dôsledky roztrieštenosti informácií a fenomén "dát v silách", ktorý bráni efektívному rozhodovaniu. Skúma sa problematika HiPPO -prístupu (Highest Paid Person's Opinion) a jeho vplyv na kvalitu manažérskych rozhodnutí v stavebných projektoch. Hodnotí sa vplyv dynamických podnikových procesov a ich rastúcej zložitosti na informačné toky a prevádzkovú efektívnosť. Uvádzajú sa konkrétné príklady toho, ako nadmerná zložitosť systémov zvyšuje náklady a znižuje flexibilitu organizácií. Osobitná pozornosť sa venuje obmedzeniam spôsobeným proprietárnymi formátmi a perspektívam používania otvorených štandardov v stavebníctve. Prezentuje sa koncepcia prechodu na softvérové ekosystémy založené na umelej inteligencii a LLM, ktoré minimalizujú nadmernú zložitosť a technické prekážky.

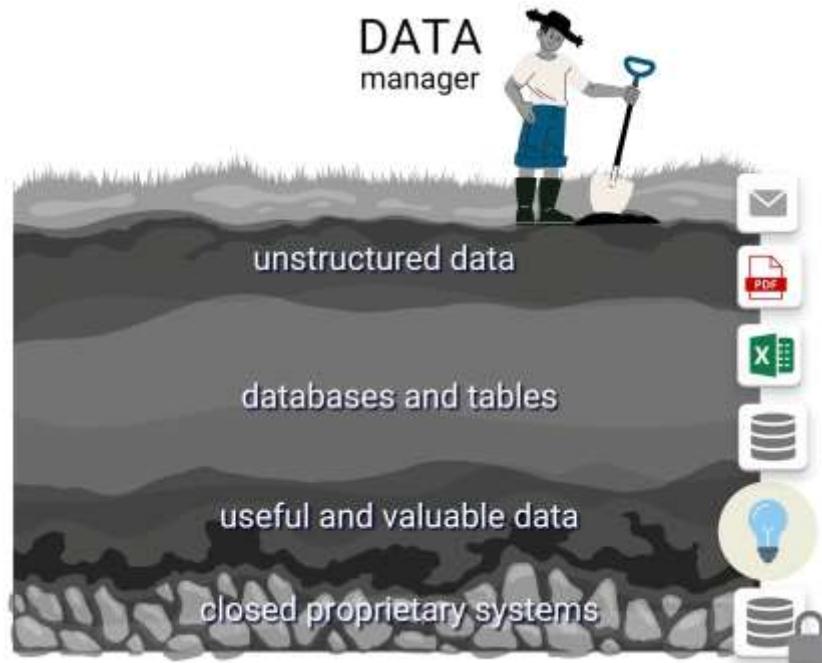
## KAPITOLA 2.1.

# ROZTRIEŠTENOSŤ ÚDAJOV A DÁTOVÉ SILÁ

### Čím viac nástrojov, tým efektívnejšie podnikanie?

Na prvý pohľad sa môže zdať, že viac digitálnych nástrojov vedie k väčšej efektívnosti. V praxi to však tak nie je. S každým novým riešením, či už ide o cloudovú službu, starší systém alebo ďalšiu správu v programe Excel, pridáva spoločnosť do svojho digitálneho prostredia ďalšiu vrstvu - vrstvu, ktorá často nie je integrovaná s ostatnými (obrázok 2.1-1).

Údaje možno prirovnáť k uhliu alebo rope: hromadia sa roky, sú zhutnené pod vrstvami chaosu, chýb, neštruktúrovaných procesov a zabudnutých formátov. Ak z nich chcú spoločnosti získať skutočne užitočné informácie, musia sa doslova predierať cez vrstvy zastaraných riešení a digitálneho šumu.



Obr. 2.1-1 Rôznorodé údaje tvoria rozdelené vrstvy - v geologických horninách systémovej zložitosti sa strácajú aj "zlaté" poznatky.

Každá nová aplikácia za sebou zanecháva stopu: súbor, tabuľku alebo celé izolované "silo" na serveri. Jedna vrstva je hlina (zastarané a zabudnuté údaje), druhá je piesok (rôznorodé tabuľky a zostavy) a tretia je žula (uzavreté proprietárne formáty, ktoré sa nedajú integrovať). Postupom času sa digitálne prostredie spoločnosti čoraz viac podobá nádrži nekontrolovaného hromadenia informácií, kde sa hodnota stráca hlboko v serveroch spoločnosti.

S každým novým projektom a každým novým systémom sa stáva zložitejšou nielen infraštruktúra, ale aj cesta k užitočným kvalitným údajom. Dostať sa k hodnotnému "kameňu" si vyžaduje hĺbkové

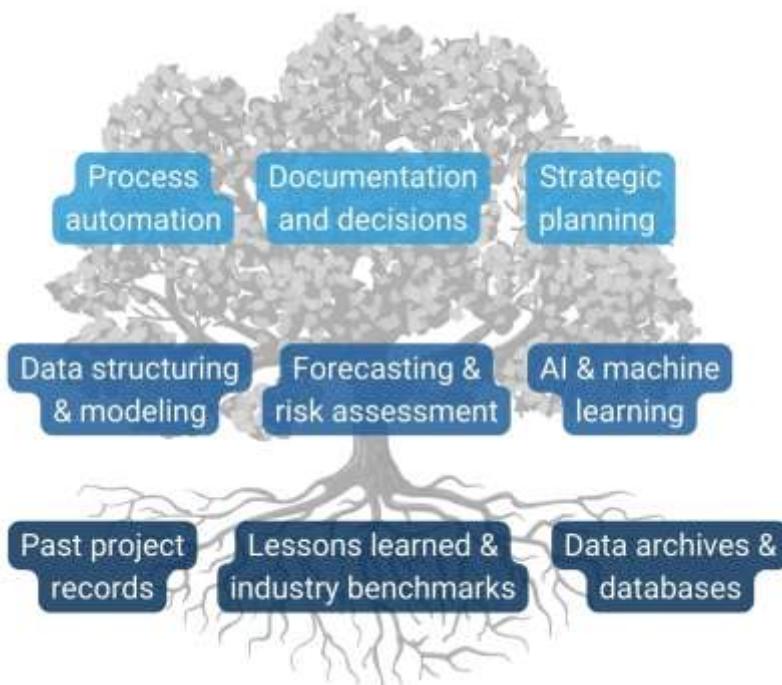
čistenie, štruktúrovanie informácií, ich "štiepenie", zoskupovanie do zmysluplných častí a získavanie strategicky dôležitých poznatkov prostredníctvom analýzy a modelovania údajov.

*Údaje sú cenná vec a vydržia dlhšie ako samotné systémy [ktoré údaje spracovávajú] [29].*

- Tim Berners-Lee, otec World Wide Webu a tvorca prvej webovej stránky

Predtým, ako sa údaje stanú "cennou vecou" a spoločne súčasťou pre rozhodovanie, musia prejsť dôkladnou prípravou. Práve správne predspracovanie mení rôznorodé údaje na štruktúrované skúsenosti, užitočný informačný humus, ktorý sa potom stáva nástrojom na prognózovanie a optimalizáciu.

Existuje mylná predstava, že na začiatku analýzy potrebujete dokonale čisté údaje, ale v praxi je schopnosť pracovať so špinavými údajmi nevyhnutnou súčasťou procesu.



Obr. 2.1-2 Údaje sú koreňovým systémom a základom podnikania, ktoré je zasa založené na rozhodovacích procesoch.

Kedže technológie neustále napredujú, aj vaša firma musí napredovať a naučiť sa vytvárať hodnotu z údajov. Tak ako ropa a uhoľné spoločnosti budujú infraštruktúru na ťažbu nerastných surovín, aj podniky sa musia naučiť, ako riadiť tok nových informácií na vlastných serveroch a získavať cenné

poznatky z nepoužívaných, neformátovaných a zastaraných údajov a premeniť ich na strategický zdroj.

Vytvorenie polí (dátových skladov) je prvým krokom. Ani tie najvýkonnejšie nástroje nevyriešia problém izolácie údajov a údajov vo viacerých formátoch, ak spoločnosti budú naďalej pracovať v izolovaných systémoch. Ak údaje existujú oddelene od seba, bez vzájomného prepojenia a zdieľania informácií, podniky čelia efektu "dátového sila". Namiesto jednotnej, konzistentnej infraštruktúry sú spoločnosti nútené vynakladať zdroje na spájanie a synchronizáciu údajov.

## Dátové silá a ich vplyv na výkonnosť spoločnosti

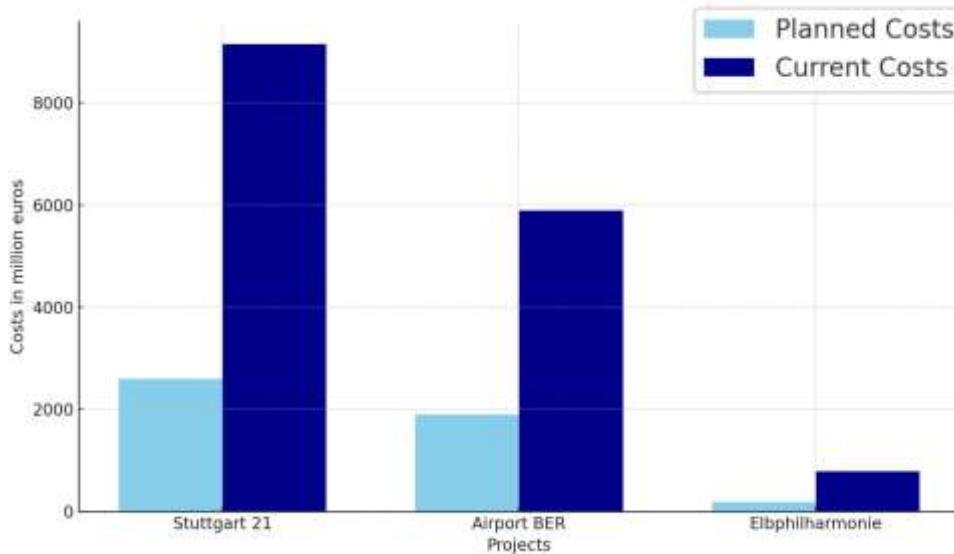
Predstavte si, že staviate sídlisko, ale každý tím má svoj vlastný projekt. Niektorí stavajú múry, iní kladú komunikácie a ďalší kladú cesty bez toho, aby sa navzájom kontrolovali. Výsledkom je, že potrubia nezodpovedajú otvorom v stenách, výtahové šachty nezodpovedajú poschodiám a cesty sa musia rozobrať a položiť nanovo.

Táto situácia nie je len hypotetickým scenárom, ale realitou mnohých moderných stavebných projektov. Vzhľadom na veľký počet generálnych a subdodávateľských firiem pracujúcich s rôznymi systémami a bez jednotného koordinačného centra sa proces mení na sériu nekonečných schvaľovaní, prepracovaní a konfliktov. To všetko vedie k výrazným oneskoreniam a niekoľkonásobným nákladom na projekt.

Klasická situácia na stavenisku je jednoduchá: debnenie je pripravené, ale dodávka výstuže nedorazila včas. Pri kontrole informácií v rôznych systémoch prebieha komunikácia zhruba takto:

- ❷ **Majster** na stavbe 20. dňa píše vedúcemu projektu: "*Dokončili sme debnenie, kde je armovacia tyč?*"
- ❷ **Projektový manažér** (PMIS) oddeleniu verejného obstarávania: - "*Debnenie je pripravené. V mojom systéme [PMIS] sa píše, že armatúra mala prísť 18. Kde sú výstuže?*"
- ❷ **Špecialista dodávateľského reťazca** (ERP): - "*Nás systém ERP hovorí, že dodávka bude 25.*"
- ❷ **Dátový inžinier** alebo IT oddelenie (zodpovedné za integrácie): - *V systéme PMIS je dátum 18., v systéme ERP je to 25. Medzi ERP a PMIS neexistuje prepojenie OrderID, takže údaje nie sú synchronizované. Toto je typický príklad informačnej medzery.*
- ❷ **Projektový manažér** generálnemu riaditeľovi - "*Dodávka armatúr mešká, stavenisko stojí a nie je jasné, kto je za to zodpovedný*".

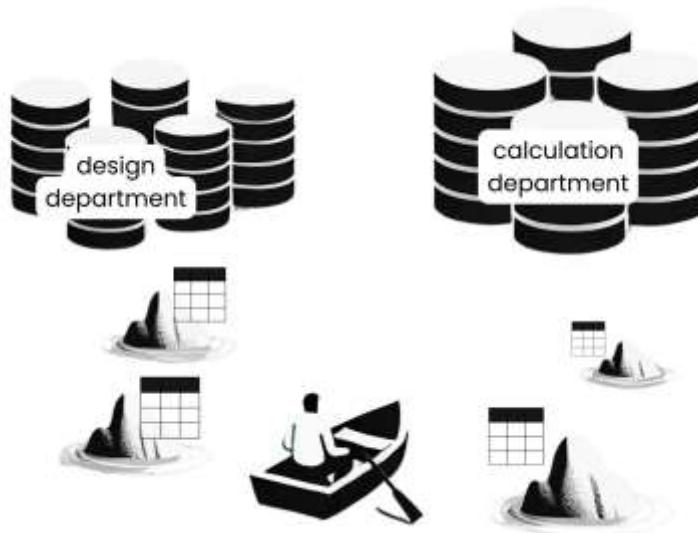
Príčinou incidentu bola izolácia údajov v rôznych systémoch. Integráciou a zjednotením zdrojov údajov, vytvorením jedného úložiska informácií a automatizáciou prostredníctvom nástrojov ETL (Apache NiFi, Airflow alebo n8n) možno odstrániť silá medzi systémami. Týmto a ďalším metódam a nástrojom sa budeme podrobne venovať v ďalších častiach knihy.



Obr. 2.1-3 Porovnanie plánovaných a skutočných nákladov veľkých infraštruktúrnych projektov v Nemecku.

To isté sa deje s podnikovými systémami: najprv sa vytvoria izolované riešenia a potom sa musia vynaložiť obrovské rozpočty na ich integráciu a harmonizáciu. Ak by sa od začiatku premysleli dátové a komunikačné modely, integrácia by vôbec nebola potrebná. Oddelené údaje vytvárajú v digitálnom svete chaos ako nekoordinovaný stavebný proces.

Podľa štúdie spoločnosti KPMG "Cue construction 4.0 2023: Time to make or break" len 36% spoločností efektívne zdieľa údaje medzi jednotlivými oddeleniami, zatiaľ čo 61% spoločností čelí vážnym problémom kvôli izolovaným dátovým "silám" [30].



Obr. 2.1-4 Roky ľažko zhromažďovaných údajov sa hromadia v izolovaných úložných "silách" s rizikom, že sa nikdy nepoužijú.

Firemné údaje sú uložené v izolovaných systémoch ako jednotlivé stromy roztrúsené po krajine. Každý z nich obsahuje cenné informácie, ale nedostatok prepojení medzi nimi bráni vytvoreniu jednotného, vzájomne prepojeného ekosystému. Toto oddelenie bráni toku údajov a obmedzuje schopnosť organizácie vidieť celý obraz. Prepojenie týchto síl je mimoriadne dlhý a zložitý proces pestovania podhubia na úrovni manažmentu, aby sa naučil prenášať jednotlivé informácie medzi systémami.

Podľa štúdie WEF z roku 2016 je jednou z hlavných prekážok digitálnej transformácie nedostatok spoločných dátových štandardov a roztriedenosť.

*Stavebný priemysel je jedným z najfragmentovanejších na svete a závisí od hladkej interakcie všetkých účastníkov hodnotového reťazca [5].*

- Svetové ekonomicke fórum 2016: Formovanie budúcnosti

Projektanti, manažéri, koordinátori a vývojári často uprednostňujú samostatnú prácu, aby sa vyhli zložitej koordinácii. Tento prirodzený sklon viedie k vytváraniu informačných "síl", v ktorých sú údaje izolované v rámci samostatných systémov. Čím viac je takýchto izolovaných systémov, tým ľahšie je dosiahnuť ich vzájomnú spoluprácu. Časom každý systém získa vlastnú databázu a špecializované podporné oddelenie manažérov (obrázok 1.2-4), čo ďalej komplikuje integráciu.



Obr. 2.1-5 Každý systém má tendenciu vytvárať vlastné unikátné silá údajov, ktoré je potrebné spracovať vhodnými nástrojmi [31].

Začarovaný kruh v podnikových systémoch vyzerá takto: spoločnosti investujú do zložitých izolovaných riešení, potom čelia vysokým nákladom na ich integráciu a vývojári, ktorí si uvedomujú zložitosť kombinovania systémov, radšej pracujú vo svojich uzavretých ekosystémoch. To všetko zvyšuje fragmentáciu IT prostredia a sťaže prechod na nové riešenia (obrázok 2.1-5). Manažéri

nakoniec kritizujú dátové silá, ale málokedy analyzujú ich príčiny a spôsoby, ako im predchádzať. Manažéri sa sťažujú na zastarané IT systémy, ale ich nahradenie si vyžaduje značné investície a málokedy prináša očakávané výsledky. Výsledkom je, že aj pokusy o boj s problémom často situáciu ešte zhoršujú.

Hlavným dôvodom odpojenia je uprednostňovanie aplikácií pred dátami. Spoločnosti najprv vyvíjajú samostatné systémy alebo kupujú hotové riešenia od dodávateľov a potom sa ich snažia zjednotiť vytvorením duplicitných a nekompatibilných úložísk a databáz.

Preknanie problému fragmentácie si vyžaduje radikálne nový prístup - uprednostnenie dát pred aplikáciami. Spoločnosti musia najprv vypracovať stratégie správy údajov a dátové modely a potom vybudovať systémy alebo nakúpiť riešenia, ktoré pracujú s jediným súborom informácií, a nie vytvárať nové prekážky.

*Vstupujeme do nového sveta, v ktorom môžu byť údaje dôležitejšie ako softvér.*

- Tim O'Reilly, generálny riaditeľ spoločnosti O'Reilly Media, Inc.

Štúdia McKinsey Global Institute "Rethinking Construction: the path to improved productivity" (2016) dokazuje, že stavebníctvo zaostáva za ostatnými sektormi v digitálnej transformácii [32]. Podľa správy môže prijatie automatizovanej správy údajov a digitálnych platform výrazne zvýšiť produktivitu a znížiť straty spojené s nekonzistentnosťou procesov. Túto potrebu digitálnej transformácie zdôrazňuje aj správa Egana (Spojené kráľovstvo, 1998) [33], ktorá zdôrazňuje klúčovú úlohu integrovaných procesov a kolaboratívneho prístupu v stavebníctve.

Výsledkom je, že zatiaľ čo v posledných 10 000 rokoch bol hlavným problémom správcov údajov nedostatok údajov, s lavinou údajov a systémov správy údajov sa používatelia a správcovia stretávajú s problémom - nadbytkom údajov, čo sťažuje vyhľadávanie právne správnych a kvalitných informácií.

Rozdielne dátové silá nevyhnutne vedú k vážnemu problému zníženej kvality údajov. Pri viacerých nezávislých systémoch môžu tie isté údaje existovať v rôznych verziach, často s protichodnými hodnotami, čo vytvára ďalšie ťažkosti pre používateľov, ktorí musia určiť, ktoré informácie sú relevantné a spoľahlivé.

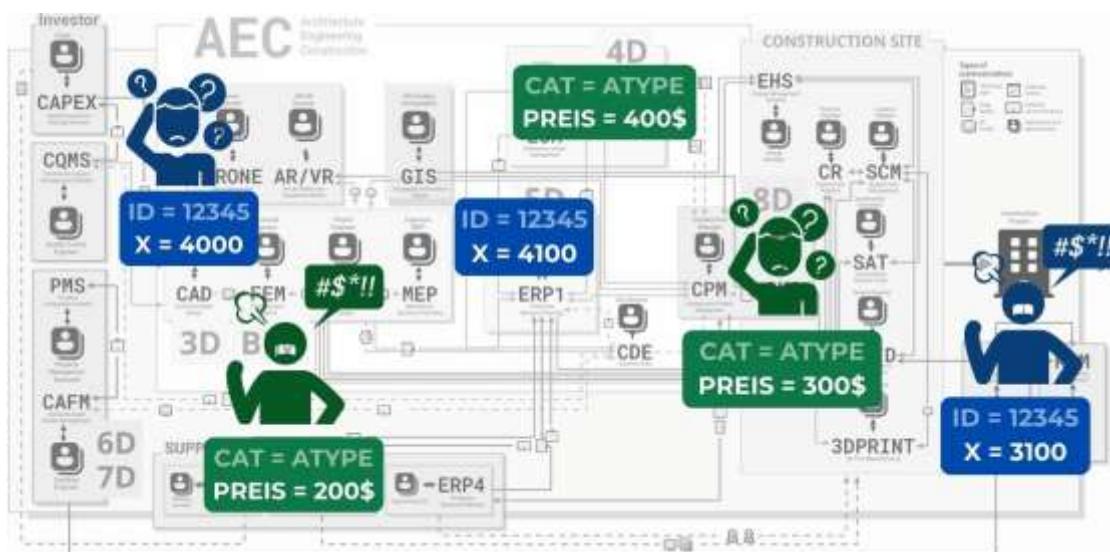
## Duplicita a nedostatočná kvalita údajov ako dôsledok nejednotnosti

Kvôli problému dátových síl musia manažéri stráviť veľa času hľadaním a zosúladovaním údajov. Aby sa spoločnosti zabezpečili proti problémom s kvalitou, vytvárajú komplexné štruktúry riadenia informácií, v ktorých je za vyhľadávanie, overovanie a zosúladovanie údajov zodpovedná vertikálna manažérov. Tento prístup však len zvyšuje byrokraciu a spomaľuje rozhodovanie. Čím viac je údajov, tým ťažšie je ich analyzovať a interpretovať, najmä ak neexistuje jednotný štandard na ich ukladanie a

spracovanie.

S množstvom softvérových aplikácií a systémov, ktoré v poslednom desaťročí rastú ako huby po daždi, sa problém s dátovými silami a nevhodnou kvalitou údajov stáva pre koncových používateľov čoraz dôležitejším. Tie isté údaje, ale s rôznymi hodnotami, sa teraz nachádzajú v rôznych systémoch a aplikáciách (obrázok 2.1-6). To vedie k ťažkostiam pre koncových používateľov pri snahe určiť, ktorá verzia údajov je relevantná a správna spomedzi mnohých dostupných. To vedie k chybám pri analýzach a v konečnom dôsledku aj pri rozhodovaní.

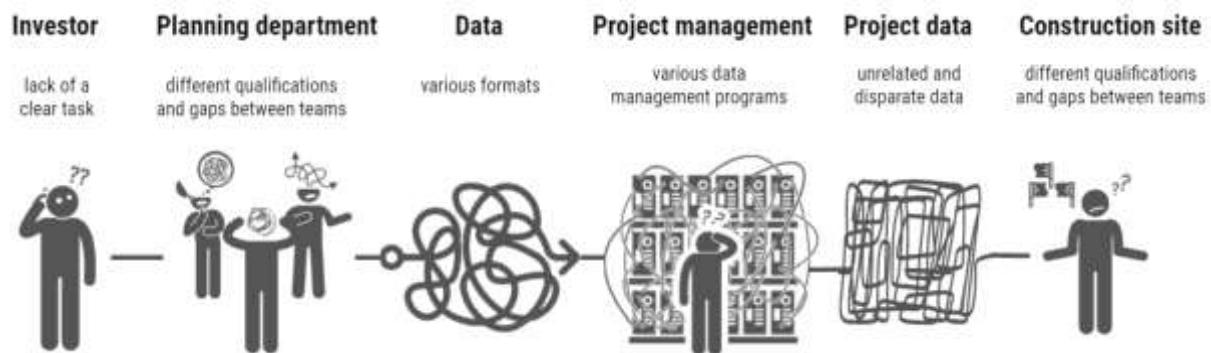
Aby sa manažéri spoločnosti poistili proti problémom s vyhľadávaním správnych údajov, vytvárajú viacúrovňovú byrokraciu manažérov overovania. Ich úlohou je vedieť rýchlo nájsť, skontrolovať a odoslať požadované údaje vo forme tabuľiek a správ, pričom sa musia orientovať v splete rôznorodých systémov.



Obr. 2.1-6 Pri snahe nájsť správne údaje musia manažéri zabezpečiť kvalitu a právnu spoľahlivosť údajov medzi rôznymi systémami.

V praxi však tento model vytvára nové zložitosti. Ak sa údaje spravujú manuálne a informácie sú rozptýlené v mnohých nesúvisiacich rozhodnutiach, každý pokus o získanie presných a aktuálnych informácií prostredníctvom pyramídy rozhodovacích orgánov (obrázok 2.1-7) sa stáva úzkym miestom - časovo náročným a náchylným na chyby.

Situáciu zhoršuje lavína digitálnych riešení. Softvérový trh je stále zaplavený novými nástrojmi, ktoré sa zdajú byť slúbné. Ale bez jasnej stratégie správy údajov sa tieto riešenia neintegrujú do jednotného systému, ale namiesto toho vytvárajú ďalšie vrstvy zložitosti a duplicity. Výsledkom je, že namiesto zjednodušenia procesov sa spoločnosti ocitajú v ešte roztrieštenejšom a chaotickejšom informačnom prostredí.



Obr. 2.1-7 Zložitosť systémov a rôznorodosť dátových formátov vedú k strate konzistentnosti v procese konštrukcie.

Všetky tieto problémy spojené so správou množstva rôznorodých riešení skôr či neskôr priviedú vedenie spoločnosti k dôležitému zisteniu: nejde o objem údajov alebo hľadanie ďalšieho "univerzálneho" nástroja na spracovanie údajov. Skutočný dôvod spočíva v kvalite údajov a v tom, ako ich organizácia vytvára, prijíma, uchováva a používa.

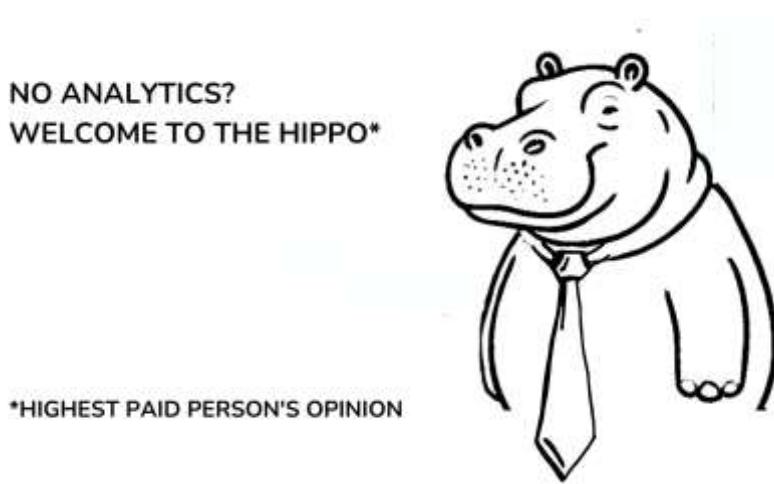
Kľúčom k udržateľnému úspechu nie je honba za novými "zázračnými" aplikáciami, ale budovanie dátovej kultúry v rámci spoločnosti. To znamená, že údaje sa považujú za strategické aktívum a kvalita, integrita a relevantnosť údajov sa stávajú prioritou na všetkých úrovniach organizácie.

Riešenie dilemy "kvalita verus kvantita" spočíva vo vytvorení jednotnej štruktúry údajov, ktorá eliminuje duplicitu, odstraňuje nekonzistentnosti a zjednocuje informačné toky. Táto štruktúra poskytuje jediný spoľahlivý zdroj údajov, na základe ktorého sa prijímanú informované, presné a včasné rozhodnutia.

V opačnom prípade, ako je to stále časté, sa spoločnosti naďalej spoliehajú skôr na subjektívne názory a intuitívne hodnotenia odborníkov HiPPO než na spoľahlivé fakty. V stavebnictve, kde odbornosť tradične zohráva významnú úlohu, je to obzvlášť viditeľné.

## HiPPO alebo nebezpečenstvo názorov pri rozhodovaní

V stavebnictve sa kľúčové rozhodnutia tradične prijímajú na základe skúseností a subjektívneho úsudku. Bez včasných a spoľahlivých údajov musia manažéri spoločnosti konať naslepo a spoliehať sa skôr na intuiciu najlepšie platených zamestnancov (HiPPO - Highest Paid Person's Opinion) než na objektívne fakty (obr. 2.1-8).



**Obrázok 2.1-8 Pri absencii analytiky podnikanie závisí od subjektívneho názoru skúsených odborníkov.**

Tento prístup môže byť opodstatnený v stabilnom a pomaly sa meniacom prostredí, ale v ére digitálnej transformácie sa stáva vážnym rizikom. Rozhodnutia založené na intuícii a odhadoch sú náchylné na skreslenie, často vychádzajú z nepodložených hypotéz a nezohľadňujú komplexný obraz, ktorý sa odráža v údajoch

To, čo sa na úrovni rozhodovania v podniku vydáva za inteligentnú diskusiu, často nie je založené na ničom konkrétnom. Úspech podniku by nemal závisieť od autority a platu odborníkov, ale od schopnosti efektívne pracovať s údajmi, identifikovať zákonitosti a prijímať informované rozhodnutia.

Je dôležité upustiť od koncepcie, že autorita alebo skúsenosť automaticky znamená, že rozhodnutie je správne. Prístup založený na údajoch mení pravidlá hry: základom rozhodovania sú teraz údaje a analytika, nie pozícia a plat. Veľké objemy údajov, strojové učenie a vizuálna analytika nám umožňujú identifikovať vzory a spoliehať sa na fakty, a nie na dohady (obr. 1.1-4).

*Bez údajov ste len ďalším človekom s názorom [34].*

- W. Edwards Deming, vedec a konzultant v oblasti riadenia

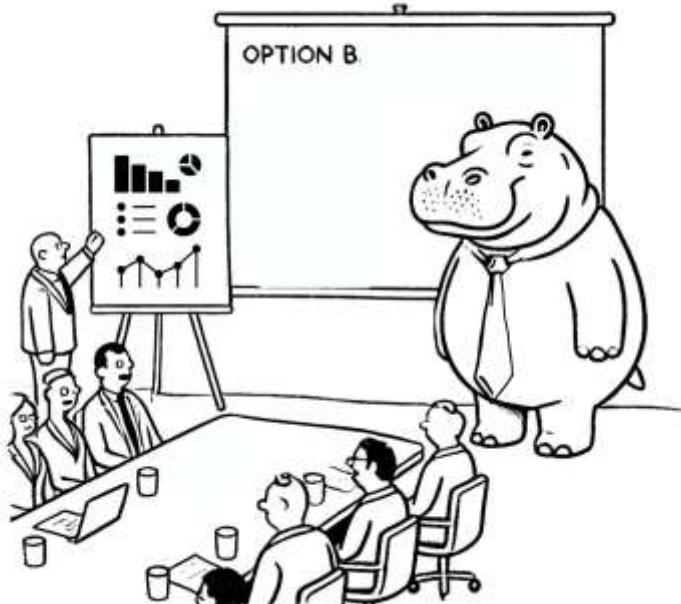
Moderné metódy správy údajov zabezpečujú aj kontinuitu znalostí v rámci spoločnosti. Jasne popísané procesy, automatizácia a systematický prístup umožňujú presunúť aj kľúčové úlohy bez straty efektívnosti.

Slepá dôvera v údaje však môže viesť aj k závažným chybám. Samotné údaje sú len súborom čísel. Bez správnej analýzy, kontextu a schopnosti identifikovať vzory nemajú žiadnu hodnotu a nemôžu riadiť procesy. Kľúč k úspechu nespočíva vo výbere medzi intuíciou HiPPO a analýzou, ale v budovaní inteligentných nástrojov, ktoré transformujú rôznorodé informácie na zvládnuteľné a informované rozhodnutia.

V digitálnom stavebnom prostredí sa rozhodujúcimi faktormi úspechu nestáva služobný postup a miesto v hierarchii, ale schopnosť reagovať, presnosť rozhodovania a efektívnosť zdrojov

Údaje sú nástroje, nie absolútne pravdy. Mali by dopĺňať ľudské myslenie, nie ho nahradzať. Napriek výhodám analytiky, údaje nemôžu úplne nahradniť ľudskú intuíciu a skúsenosť. Ich úlohou je pomáhať prijímať presnejšie a informovanejšie rozhodnutia.

Konkurenčná výhoda sa nedosiahne len dodržiavaním noriem, ale aj tým, že dokážeme predbehnuť konkurenciu v efektívnom využívaní zdrojov, ktoré sú pre všetkých rovnaké. V budúcnosti sa zručnosti v oblasti údajov stanú rovnako dôležitými ako kedysi gramotnosť alebo matematické zručnosti. Odborníci, ktorí dokážu analyzovať a interpretovať údaje, budú schopní prijímať presnejšie rozhodnutia a vytlačia tých, ktorí sa spoliehajú len na osobné skúsenosti (obrázok 2.1-9).



Obr. 2.1-9 Rozhodnutia by mali vychádzať z objektívnej analýzy, nie z názoru najlepšie plateného zamestnanca.

Manažéri, špecialisti a inžinieri budú pôsobiť ako dátoví analytici, ktorí budú skúmať štruktúru, dynamiku a klúčové ukazovatele projektov. Ľudské zdroje sa stanú prvkami systému, ktoré si budú vyžadovať flexibilné prispôsobenie na základe údajov s cieľom maximalizovať efektívnosť.

*Chyby pri použití nedostatočných údajov sú oveľa menšie ako pri použití žiadnych údajov [35].*

- Charles Babbage, vynálezca prvého analytického počítacieho stroja

Vznik veľkých dát a zavedenie LLM (Large Language Models) radikálne zmenili nielen spôsob analýzy, ale aj samotnú podstatu rozhodovania. Kým predtým sa pozornosť sústredovala na kauzalitu (prečo sa niečo stalo - diagnostická analytika) (obr. 1.1-4), dnes sa do popredia dostáva schopnosť predvídať budúnosť (prediktívna analytika) a v budúnosti aj preskriptívna analytika, kde strojové učenie a umelá inteligencia navrhujú najlepšiu voľbu v procese rozhodovania.

Podľa novej štúdie spoločnosti SAP™ "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" (Nová štúdia zistila, že takmer polovica vedúcich pracovníkov dôveruje umelej inteligencii viac ako sebe) z roku 2025 [36] by 44% vrcholových manažérov bolo ochotných zmeniť svoje predchádzajúce rozhodnutie na základe rady umelej inteligencie, a 38% by dôverovalo umelej inteligencii, aby v ich mene prijímal obchodné rozhodnutia. Pritom 74% vedúcich pracovníkov uviedlo, že dôverujú radám AI viac ako svojim priateľom a rodine, a 55% pracuje v spoločnostiach, kde poznatky získané pomocou AI nahrádzajú alebo často obchádzajú tradičné metódy rozhodovania - najmä v organizáciách s ročnými príjmami nad 5 miliárd USD. Okrem toho 48% respondentov používa generatívne nástroje AI na dennej báze, vrátane 15%, ktorí ich používajú niekoľkokrát denne.

S rozvojom systémov LLM a automatizovanej správy údajov vzniká nová výzva: ako efektívne využívať informácie bez toho, aby sa stratila ich hodnota v chôose nekompatibilných formátov a heterogénnych zdrojov, ktorý dopĺňa rastúca zložitosť a dynamika podnikových procesov.

## **Neustály nárast zložitosti a dynamiky obchodných procesov**

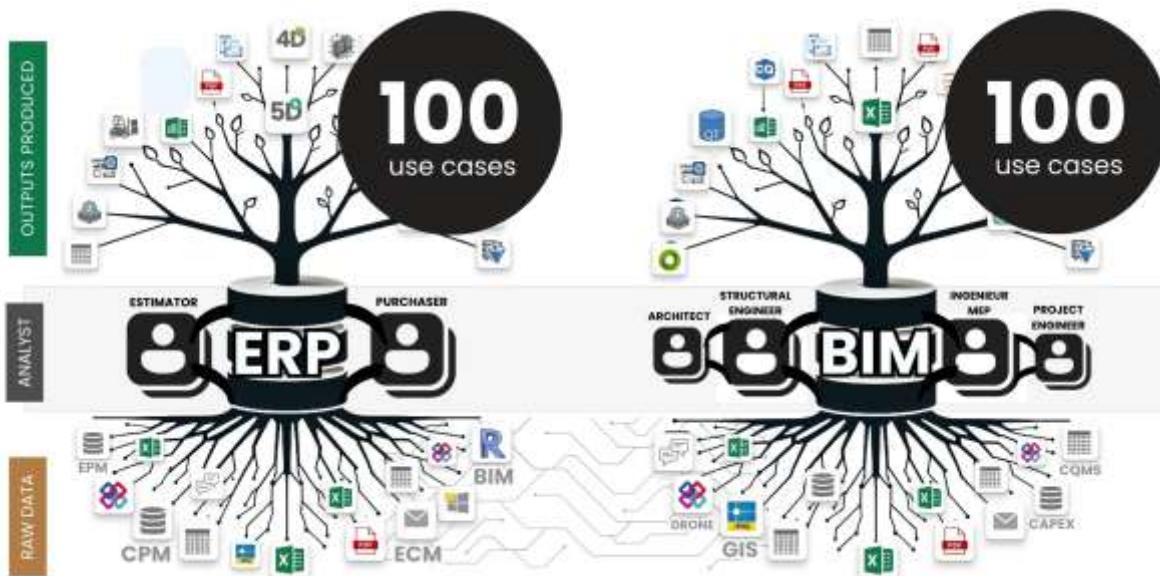
Stavebníctvo dnes čelí vážnym výzvam v oblasti správy údajov a procesov. Hlavnými problémami sú izolované informačné systémy, nadmerná byrokracia a nedostatočná integrácia digitálnych nástrojov. Tieto výzvy sa stupňujú, pretože samotné obchodné procesy sú čoraz zložitejšie - pod vplyvom technológií, meniacich sa požiadaviek zákazníkov a vyvíjajúcich sa predpisov.

Jedinečnosť stavebných projektov vyplýva nielen z ich technických zvláštností, ale aj z rozdielov v národných normánoch a regulačných požiadavkách v rôznych krajinách (obr. 4.2-10, obr. 5.1-7). To si vyžaduje flexibilný, individuálny prístup ku každému projektu, ktorý je v rámci tradičných modulárnych systémov riadenia ľahko realizovateľný. Vzhľadom na zložitosť procesov a veľké množstvo údajov sa mnohé spoločnosti obracajú na dodávateľov, ktorí ponúkajú špecializované riešenia. Trh je však preťažený - mnoho začínajúcich firiem ponúka podobné produkty, ktoré sa zameriavajú na úzke úlohy. V dôsledku toho sa často stráca holistiký prístup k správe údajov.

Rozhodujúcim faktorom konkurencieschopnosti sa stáva prispôsobovanie sa neustálemu prílevu nových technológií a požiadaviek trhu. Existujúce proprietárne aplikácie a modulárne systémy však majú nízku prispôsobivosť - akékoľvek zmeny si často vyžadujú zdíľavé a nákladné revízie zo strany vývojárov, ktorí nie vždy rozumejú špecifikám stavebných procesov.

Spoločnosti sa stávajú rukojemníkmi technologického oneskorenia a čakajú na nové aktualizácie namiesto toho, aby urýchlene zaviedli inovatívne integrované prístupy. Výsledkom je, že vnútorná

štruktúra stavebných organizácií je často zložitý ekosystém vzájomne prepojených hierarchických a často uzavretých systémov koordinovaných prostredníctvom viacúrovňovej siete manažérov (obrázok 2.1-10).

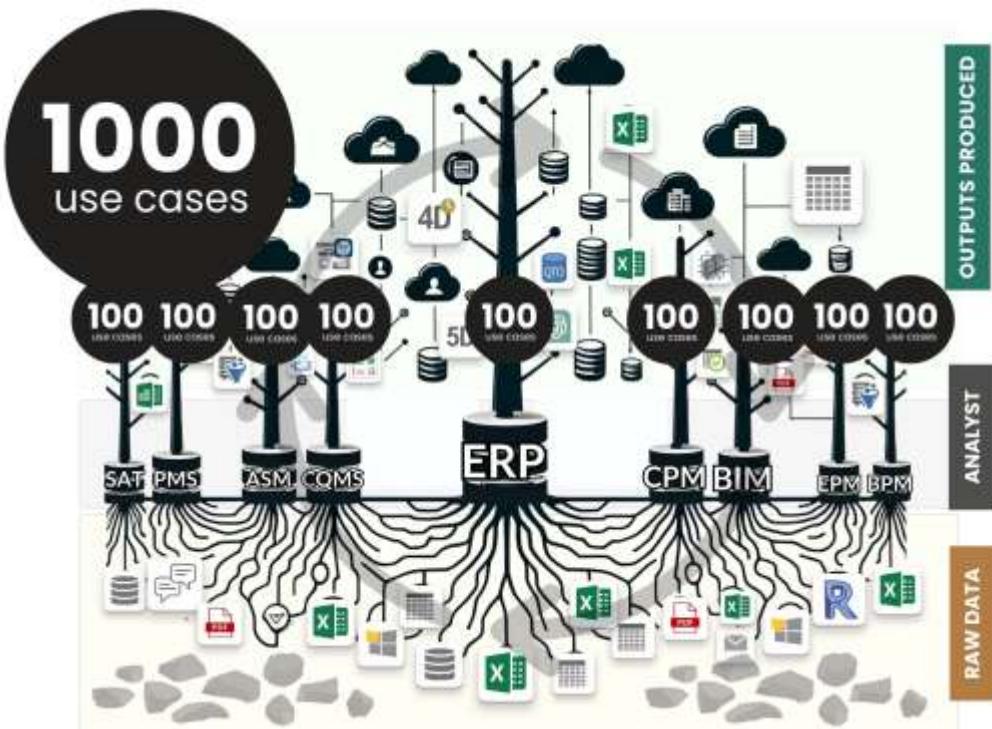


Obr. 2.1-10 Podniky sa skladajú zo vzájomne prepojených systémov, ktorých vzájomné prepojenie vytvára procesy, ktoré si vyžadujú automatizáciu.

Podľa prieskumu, ktorý v roku 2021 uskutočnili Kanadská stavebná asociácia a spoločnosť KPMG Canada [37], sa len 25% spoločností domnieva, že sú v porovnaní s konkurenciou vo významnej alebo odlišnej pozícii, pokial' ide o zavádzanie technológií alebo digitálnych riešení. Iba 23% respondentov uviedlo, že ich riešenia sú významne alebo vo veľkej mieri založené na údajoch. Väčšina respondentov prieskumu zároveň charakterizovala svoje využívanie celého radu ďalších technológií ako čisto experimentálne alebo priznala, že ich vôbec nevyužíva.

Táto neochota podieľať sa na technologických experimentoch je zrejmá najmä pri veľkých infraštrukturálnych projektoch, kde chyby môžu stáť milióny dolárov. Dokonca aj tie najpokročilejšie technológie - digitálne dvojčatá, prediktívna analytika - často narážajú na odpor nie kvôli svojej účinnosti, ale kvôli nedostatočnej spôsobilosti overenej v reálnych projektoch.

Podľa správy Svetového ekonomického fóra (WEF) "Formovanie budúcnosti stavebníctva" [5] naráža zavádzanie nových technológií v stavebníctve nielen na technické problémy, ale aj na psychologické prekážky zo strany zákazníkov. [5], zavádzanie nových technológií v stavebníctve naráža nielen na technické ťažkosti, ale aj na psychologické bariéry zo strany zákazníkov. Mnohí zákazníci sa obávajú, že používanie moderných riešení urobí z ich projektov experimentálne pracovisko a urobí z nich "pokusných králikov" a nepredvídateľné dôsledky môžu viesť k dodatočným nákladom a rizikám.



Obr. 2.1-11 Pre každý prípad použitia údajov ponúka trh s riešeniami aplikácie na optimalizáciu a automatizáciu procesov.

Stavebníctvo je veľmi rozmanité: rôzne projekty majú rôzne požiadavky, regionálne špecifika, zákonné klasifikačné predpisy (obr. 4.2-10), výpočtové normy (obr. 5.1-7) atď. Preto je prakticky nemožné vytvoriť vlastnú univerzálnu aplikáciu alebo systém, ktorý by dokonale vyhovoval všetkým týmto požiadavkám a špecifikám projektu.

V snahe vyrovnať sa s rastúcou zložitosťou systémov a závislosťou od dodávateľov softvéru sa čoraz viac uvedomuje, že klúcom k efektívnej správe údajov je nielen otvorenosť a štandardizácia, ale aj zjednodušenie samotnej architektúry procesov. Rastúca zložitosť a dynamika podnikových procesov si vyžaduje nové prístupy, pri ktorých sa priorita presúva z akumulácie údajov na ich štruktúrovanie a organizáciu. Práve tento posun bude ďalším krokom vo vývoji stavebníctva, ktorý bude znamenať koniec éry dominancie dodávateľov softvéru a začiatok éry zmysluplnnej organizácie informácií.

Uvedomenie si obmedzení univerzálnych riešení a zraniteľnosti voči rastúcej zložitosti vedie k zmene priorít z uzavretých platform a hromadenia údajov na transparentnosť, prispôsobivosť a štruktúrované spracovanie informácií. Tento posun v myslení odráža širšie zmeny v globálnej ekonomike a technológiách, opísané cez prizmu tzv. priemyselných revolúcií. Aby sme pochopili, kam stavebníctvo smeruje a aké je jeho budúce smerovanie, je potrebné zvážiť miesto tohto odvetvia v kontexte štrvej a piatej priemyselnej revolúcie - od automatizácie a digitalizácie po personalizáciu, otvorené štandardy a dátový model založený na službách.

## Štvrtá priemyselná revolúcia (Priemysel 4.0) a piata priemyselná revolúcia (Priemysel 5.0) v stavebníctve

Technologické a ekonomicke štádiá sú teoretické koncepty používané na opis a analýzu vývoja spoločnosti a hospodárstva v rôznych štádiach vývoja. Rôzni výskumníci a odborníci ich môžu interpretovať rôzne.

- **Štvrtá priemyselná revolúcia (4IR alebo Industry 4.0)** súvisí s informačnými technológiami, automatizáciou, digitalizáciou a globalizáciou. Jedným z jej kľúčových prvkov je vytváranie vlastných softvérových riešení, t. j. špecializovaných digitálnych produktov určených pre konkrétné úlohy a podniky. Tieto riešenia sa často stávajú dôležitou súčasťou IT infraštruktúry, ale bez dodatočných úprav sú slabo škálovateľné.
- **Piata priemyselná revolúcia (5IR)** je v súčasnosti v skoršej fáze koncepcionalizácie a vývoja ako 4IR. Medzi jej hlavné zásady patrí väčšia personalizácia výrobkov a služieb. 5IR je smerovanie k prispôsobivejšej, flexibilnejšej a personalizovanejšej hospodárskej činnosti so zameraním na personalizáciu, poradenstvo a modely orientované na služby. Kľúčovým aspektom piatého hospodárskeho spôsobu je využívanie údajov na rozhodovanie, ktoré je prakticky nemožné bez využívania otvorených údajov a otvorených nástrojov (obrázok 2.1-12).



Obr. 2.1-12 Štvrtý vzor sa zameriava na riešenia, zatiaľ čo piaty vzor sa zameriava na personalizáciu a údaje.

Vytvorenie aplikácie pre spoločnosti v stavebníctve na použitie v desiatich alebo sto organizáciách nezarúčuje jej úspešné rozšírenie do ďalších spoločností, regiónov alebo krajín bez výrazných úprav a vylepšení. Pravdepodobnosť úspešného škálovania takýchto riešení zostáva nízka, pretože každá organizácia má jedinečné procesy, požiadavky a podmienky, ktoré si môžu vyžadovať individuálne úpravy.

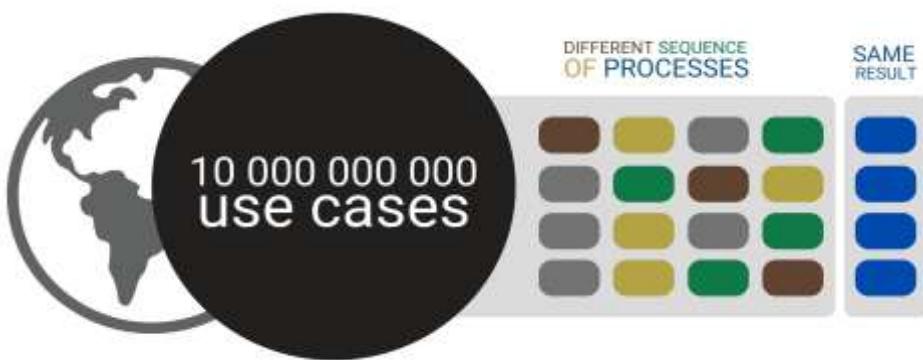
Je dôležité si uvedomiť, že úspešná integrácia technologických riešení už dnes predpokladá individuálny prístup ku každému procesu, projektu a spoločnosti. To znamená, že aj po vytvorení univerzálneho rámca, nástroja alebo programu bude potrebné jeho podrobné prispôsobenie a prispôsobenie tak, aby spĺňal jedinečné požiadavky a podmienky každej konkrétnej spoločnosti a

projektu.

Podľa správy PwC "Dekódovanie piatej priemyselnej revolúcie" [38] sa v tomto roku približne 50% vrcholových manažérov v rôznych odvetviach spolieha na integráciu pokročilých technológií a ľudských odborných znalostí. Tento prístup im umožňuje rýchlo sa prispôsobiť zmenám v dizajne výrobkov alebo požiadavkám zákazníkov a vytvoriť tak personalizovanú výrobu.

Každý proces si vyžaduje vývoj jedinečnej funkcie alebo aplikácie, čo vzhľadom na veľkosť globálneho stavebného priemyslu a rôznorodosť projektov vedie k existencii obrovského množstva obchodných prípadov, ktoré zakaždým predstavujú jedinečnú logiku potrubia (obr. 2.1-13). Každý takýto prípad má svoje vlastné špecifika a vyžaduje si individuálny prístup. Rôznorodosť možných riešení tohto istého analytického problému v kontexte rôznych prístupov sa budeme podrobnejšie zaoberať v kapitole venovanej strojovému učeniu a analýze súboru údajov Titanic (obr. 9.2-9).

**Pipeline** v kontexte digitálnych procesov je postupnosť činností, procesov a nástrojov, ktoré umožňujú automatizovaný alebo štruktúrovaný tok údajov a práce v rôznych fázach životného cyklu projektu.



Obr. 2.1-13 Individuálnosť a variabilita obchodných prípadov znemožňuje pokusy o vytvorenie škálovateľných uzavretých platform a nástrojov.

Náš život sa už v mnohých ohľadoch zmenil pod vplyvom digitálnej transformácie a dnes môžeme hovoriť o nástupe novej etapy v hospodárskom rozvoji stavebníctva. V tejto "novej ekonomike" bude hospodárska súťaž organizovaná podľa iných pravidiel: kto dokáže efektívne premeniť verejné znalosti a otvorené údaje na žiadané produkty a služby, získava v podmienkach piatej priemyselnej revolúcie klúčovú výhodu.

Ako uvádzá ekonómka Kate Maskus v knihe "Private Rights and Public Problems: The Global Intellectual Property Economy in the 21st Century" [39]. 2012 [39], "žijeme v globálnej znalostnej ekonomike a budúcnosť patrí tým, ktorí vedia, ako premeniť vedecké objavy na tovar".

Prechod na piaty ekonomický režim znamená zmenu zamerania z uzavretých IT riešení na otvorené štandardy a platformy. Spoločnosti začnú upúšťať od tradičných softvérových produktov v prospech

modelov orientovaných na služby, kde hlavným aktívom budú údaje, a nie proprietárne technológie.

Štúdia Harvard Business School 2024 [40] poukazuje na obrovskú ekonomickú hodnotu otvoreného softvéru (Open Source Software, OSS). Podľa tejto štúdie je OSS prítomný v 96% všetkých softvérových kódov a niektoré komerčné softvéry pozostávajú z 99,9% komponentov OSS. Bez OSS by spoločnosti vynaložili na softvér 3,5-krát viac prostriedkov.

Budovanie podnikových ekosystémov bude v súlade s globálnymi trendmi postupne prechádzať na piatu ekonomickú paradigmu, v ktorej sa analytické a poradenské služby začínané na údaje stanú vysšou prioritou ako izolované uzavreté riešenia s pevne definovanými scenármami použitia.

Éra digitalizácie zmení pomer síl v odvetví: namiesto spoliehania sa na riešenia od dodávateľov budú spoločnosti zakladať svoju konkurencieschopnosť na schopnosti efektívne využívať údaje. V dôsledku toho sa stavebný priemysel posunie od starých rigidných systémov k flexibilným, adaptívnym ekosystémom, v ktorých budú základom riadenia projektov otvorené štandardy a interoperabilné nástroje. Koniec éry dominancie dodávateľov aplikácií vytvorí nové prostredie, v ktorom hodnotu nebude definovať vlastníctvo uzavretého zdrojového kódu a špecializovaných konektorov, ale schopnosť premeniť údaje na strategickú výhodu.



## KAPITOLA 2.2.

### PREMENA CHAOSU NA PORIADOK A ZNÍŽENIE ZLOŽITOSTI

#### Zbytočný kód a uzavreté systémy ako prekážka zvyšovania produktivity

V uplynulých desaťročiach boli technologické zmeny v oblasti IT poháňané predovšetkým dodávateľmi softvéru. Tí určovali smer vývoja a určovali, ktoré technológie by mali spoločnosti prijať a ktoré by mali zostať v úzadí. V ére prechodu od izolovaných riešení k centralizovaným databázam a integrovaným systémom dodávateľia presadzovali licencované produkty, ktoré poskytovali kontrolu nad prístupom a škálovateľnosťou. Neskôr, s príchodom cloudových technológií a modelov softvéru ako služby (SaaS), sa táto kontrola zmenila na model predplatného, čím sa používateľia upevnilo ako lojalní zákazníci digitálnych služieb.

Tento prístup viedol k paradoxu: napriek bezprecedentnému objemu vytvoreného programového kódu sa v skutočnosti používa len malá časť z neho. Kódov je možno stokrát alebo tisíckrát viac, ako je potrebné, pretože tie isté podnikové procesy sú popísané a duplikované v desiatkach alebo stovkách programov rôznymi spôsobmi, a to aj v rámci tej istej spoločnosti. Zároveň sa už zaplatili náklady na vývoj, ktoré sú nenávratné. Napriek tomu priemysel pokračuje v reprodukcii tohto cyklu a vytvára nové produkty s minimálnou pridanou hodnotou pre konečného používateľa, častejšie pod tlakom očakávaní trhu ako skutočných potrieb.

Podľa príručky Defence Acquisition University (DAU) Software Development Cost Estimating Guide [41] sa náklady na vývoj softvéru môžu výrazne lísiť v závislosti od viacerých faktorov vrátane zložitosti systému a zvolenej technológie. Historicky sa náklady na vývoj v roku 2008 pohybovali okolo 100 USD na riadok zdrojového kódu (SLOC), zatiaľ čo náklady na údržbu sa môžu vyšplhať až na 4 000 USD na SLOC.

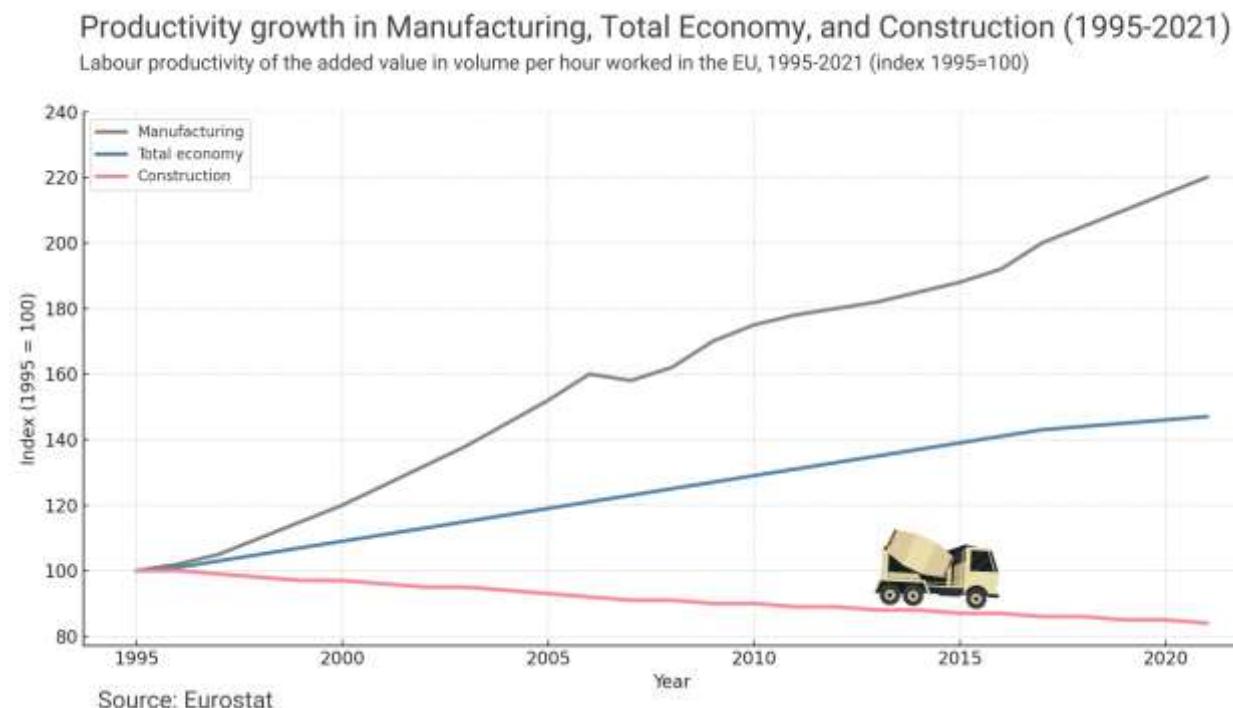
Len jedna zo súčasti CAD aplikácií - geometrické jadro - môže mať desiatky miliónov riadkov kódu (obr. 6.1-5). Podobná situácia je aj v systémoch ERP (obr. 5.4-4), k diskusii o ktorých zložitosti sa vrátime v piatej časti knihy. Pri bližšom pohľade však zistíme, že veľká časť tohto kódu neprináša pridanú hodnotu, ale funguje len ako "poštár" - mechanicky presúva údaje medzi databázou, API, používateľským rozhraním a ďalšími tabuľkami v systéme. Napriek populárному mýtu o rozhodujúcim význame takzvanej biznis logiky je krutá realita oveľa prozaickejšia: moderné kódové základne sú plné zastaraných šablónových blokov (legacy code), ktorých jediným účelom je zabezpečiť prenos údajov medzi tabuľkami a komponentmi bez vplyvu na rozhodovanie alebo efektivitu podnikania.

V dôsledku toho sa uzavreté riešenia, ktoré spracúvajú údaje z rôznych zdrojov, nevyhnutne menia na zmätočné "špagetové ekosystémy". Tieto zložité, navzájom prepletené systémy dokáže zvládnúť len armáda manažérov pracujúcich v poloprázdnom režime. Takáto organizácia správy údajov je nielen neefektívna z hľadiska zdrojov, ale vytvára aj kritické slabiny v obchodných procesoch, čím sa spoločnosť stáva závislou od úzkeho okruhu odborníkov, ktorí rozumejú fungovaniu tohto technologického bludiska.

Neustály nárast množstva kódu, počtu aplikácií a rastúca zložitosť konceptov ponúkanych dodávateľmi viedli k prirodzenému výsledku - nárastu zložitosti ekosystému IT v stavebníctve. To spôsobilo, že praktická implementácia digitalizácie prostredníctvom zvyšovania počtu aplikácií v odvetví je neefektívna. Softvérové produkty vytvorené bez náležitého zohľadnenia potrieb používateľov si často vyžadujú značné zdroje na implementáciu a podporu, ale neprinášajú očakávanú návratnosť.

Podľa štúdie spoločnosti McKinsey "Zvyšovanie produktivity v stavebníctve" [42], za posledné dve desaťročia bol celosvetový rast produktivity práce v stavebníctve v priemere len 1% ročne v porovnaní s rastom 2,8% v celom svetovom hospodárstve a 3,6% v priemyselnej výrobe. V Spojených štátach sa produktivita práce v stavebníctve na pracovníka od 60. rokov 20. storočia znížila o polovicu [43].

Rastúca zložitosť systémov, izolácia a uzavretosť údajov zhoršili komunikáciu medzi odborníkmi, čím sa stavebníctvo stalo jedným z najmenej efektívnych odvetví (obrázok 2.2-1). na 22 biliónov USD do roku 2040, čo si bude vyžadovať výrazné zvýšenie efektívnosti.



Obr. 2.2-1 Uzavreté a komplexné údaje a v dôsledku toho slabá komunikácia medzi odborníkmi viedli k tomu, že stavebníctvo patrí medzi najmenej efektívne odvetvia hospodárstva (na základe [44], [45]).

Ako sa zdôrazňuje v štúdiu spoločnosti McKinsey (2024) "Zabezpečenie produktivity v stavebníctve už nie je voliteľné", vzhľadom na rastúci nedostatok zdrojov a snahu odvetvia zdvojnásobiť tempo rastu si stavebníctvo už nemôže dovoliť zostať na súčasnej úrovni produktivity [44]. Predpokladá sa, že globálne náklady na stavebníctvo sa zvýsia z 13 biliónov USD v roku 2023 na oveľa vyššiu úroveň do konca desaťročia, čím sa otázka efektívnosti stáva nielen relevantnou, ale aj kritickou.

Jedným z kľúčových spôsobov zvýšenia efektívnosti bude nevyhnutné zjednotenie a zjednodušenie štruktúr aplikácií a architektúr dátových ekosystémov. Tento prístup k racionalizácii odstráni nadbytočné vrstvy abstrakcie a zbytočnú zložitosť, ktoré sa v podnikových systémoch nahromadili v priebehu rokov.

## **Od silosov k jednotnému dátovému skladu**

Čím viac údajov organizácia zhromažďuje, tým ľahšie je z nich získať skutočnú hodnotu. Vzhľadom na roztrieštenosť ukladania informácií v izolovaných silách sú obchodné procesy moderných spoločností ako stavitelia, ktorí sa snažia postaviť mrakodrap z materiálov uložených v tisícoch rôznych skladov. Nadbytok informácií nielenže stáže prístup k právne relevantným informáciám, ale aj spomaľuje rozhodovanie: každý krok je potrebné opakovane kontrolovať a potvrdzovať.

Každá úloha alebo proces je pevne pripojený k samostatnej tabuľke alebo databáze a výmena údajov medzi systémami si vyžaduje zložité integrácie. Chyby a nekonzistentnosti v jednom systéme môžu spôsobiť reťazové zlyhania v ostatných. Nesprávne hodnoty, oneskorené aktualizácie a duplicitné informácie nútia zamestnancov tráviť značný čas ručným zosúlaďovaním a porovnávaním údajov. V dôsledku toho organizácia trávi viac času riešením dôsledkov roztrieštenosti než vývojom a optimalizáciou procesov

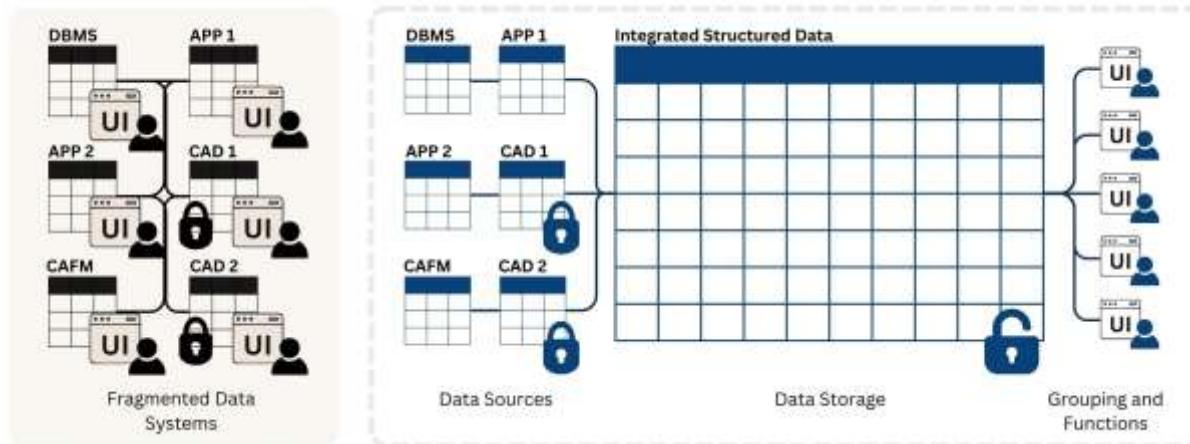
Tento problém je univerzálny: niektoré spoločnosti naďalej bojujú s chaosom, zatiaľ čo iné nachádzajú riešenie v integrácii - presúvajú informačné toky do centralizovaného úložného systému. Predstavte si ho ako jednu veľkú tabuľku, do ktorej môžete ukladať všetky entity súvisiace s úlohami, projektmi a objektmi. Namiesto desiatok rôznorodých tabuľiek a formátov vzniká jedno ucelené úložisko (obrázok 2.2-2), ktoré umožňuje:

- minimalizovať stratu údajov;
- odstrániť potrebu neustálej harmonizácie informácií;
- zlepšiť dostupnosť a kvalitu údajov;
- zjednodušenie analytického spracovania a strojového učenia

Zosúladenie údajov so spoločným štandardom znamená, že bez ohľadu na zdroj sa informácie prevedú do jednotného a strojovo čitateľného formátu. Takáto organizácia údajov umožňuje kontrolovať ich integritu, analyzovať ich v reálnom čase a pohotovo ich využívať na prijímanie manažérskych rozhodnutí.

Koncepcii integrovaných systémov ukladania dát a ich využitiu v analytike a strojovom učení sa budeme podrobnejšie venovať v kapitole "Ukladanie veľkých objemov dát a strojové učenie". Témami modelovania a štruktúrovania údajov sa budú podrobne zaoberať kapitoly "Transformácia údajov do štruktúrovanej podoby" a "Ako štandardy menia hru: od náhodných súborov k prepracovanému

dátovému modelu".



Obr. 2.2-2 Dátová integrácia odstraňuje silá, zlepšuje dostupnosť informácií a optimalizuje podnikové procesy.

Po štruktúrovaní a zlúčení údajov je ďalším logickým krokom ich validácia. Vďaka jedinému integrovanému úložisku sa tento proces výrazne zjednoduší: už žiadne viaceré nekonzistentné schémy, duplicitné štruktúry a zložité vzťahy medzi tabuľkami. Všetky informácie sú zosúladené s jediným dátovým modelom, čím sa eliminujú vnútorné nekonzistentnosti a urýchľuje sa proces validácie. Validácia a zabezpečenie kvality údajov sú základnými aspektmi všetkých obchodných procesov a podrobnejšie sa im budeme venovať v príslušných kapitolách knihy.

V záverečnej fáze sa údaje zoskupia, filtrujú a analyzujú. Aplikujú sa na ne rôzne funkcie: agregácia (scítanie, násobenie), výpočty medzi tabuľkami, stĺpcami alebo riadkami (obr. 2.2-4). Práca s údajmi sa stáva postupnosťou krov: zber, štruktúrovanie, overovanie, transformácia, analytické spracovanie a odovzdanie do konečných aplikácií, kde sa informácie používajú na riešenie praktických problémov. Viac o vytváraní takýchto scenárov, automatizácii krov a budovaní tokov spracovania si povieme v kapitolách o ETL -procesoch a prístupe k dátovým potrubiam.

Digitálna transformácia teda nie je len o zjednodušení práce s informáciami. Ide o odstránenie nadmernej zložitosti pri správe údajov, prechod od chaosu k predvídateľnosti, od viacerých systémov k riadenému procesu. Čím nižšia je zložitosť architektúry, tým menej kódu je potrebné na jej podporu. A v budúcnosti môže kód ako taký úplne zmiznúť a ustúpiť inteligentným agentom, ktorí nezávisle analyzujú, systematizujú a transformujú údaje.

## Integrované úložné systémy umožňujú prechod na agentov s umelou inteligenciou

Čím sú údaje a systémy menej zložité, tým menej kódu musíte napísat a udržiavať. A najjednoduchší spôsob, ako ušetriť vývoj, je úplne sa zbaviť kódu a nahradíť ho údajmi. Keď sa vývoj kódu aplikácií presunie z kódu na dátové modely, nevyhnutne dochádza k posunu smerom k dátovo orientovanému prístupu (data-driven), pretože za týmito koncepciami je úplne iný spôsob myšlenia.

Ked' sa človek rozhodne pracovať s údajmi v centre, začne ich úlohu vnímať inak. Údaje už nie sú len "surovinou" pre aplikácie - teraz sú základom, okolo ktorého sa buduje architektúra, logika a interakcia.

Tradičný prístup k správe údajov sa zvyčajne začína na úrovni aplikácie a v konštrukcii pripomína ťažkopádny byrokratický systém: viacúrovňové schvaľovanie, manuálne kontroly, nekonečné verzie dokumentov prostredníctvom príslušných softvérových produktov. S rozvojom digitálnych technológií bude čoraz viac spoločností nútených prejsť na princíp minimalizmu - ukladať a používať len to, čo je skutočne potrebné a čo sa bude používať.

Logiku minimalizácie prevzali predajcovia. V záujme zjednodušenia ukladania a spracovania údajov sa práca používateľov presúva z offline aplikácií a nástrojov do cloudových služieb a tzv. riešení SaaS.

Koncepcia SaaS (softvér ako služba alebo "softvér ako služba") je jedným z kľúčových trendov v moderných IT infraštruktúrach, ktorý umožňuje používateľom prístup k aplikáciám prostredníctvom internetu bez toho, aby museli inštalovať a udržiavať softvér na vlastných počítačoch.

Na jednej strane SaaS uľahčilo škálovanie, riadenie verzií a znížilo náklady na podporu a údržbu, ale na druhej strane okrem závislosti od logiky konkrétnej aplikácie spôsobilo aj úplnú závislosť používateľa od cloudovej infraštruktúry poskytovateľa. V prípade výpadku služby môže dôjsť k dočasnému alebo dokonca trvalému zablokovaniu prístupu k údajom a obchodným procesom. Okrem toho sú všetky údaje používateľa pri práci s aplikáciami SaaS uložené na serveroch poskytovateľa, čo vytvára bezpečnostné riziká a riziká súvisiace s dodržiavaním právnych predpisov. Zmeny tarif alebo podmienok používania môžu mať za následok aj zvýšené náklady alebo potrebu urgentnej migrácie.

Vývoj umelej inteligencie, LLM -agentov a prístup zameraný na údaje spochybnil budúcnosť aplikácií v ich tradičnej podobe a realizáciu SaaS. Kým predtým sa od aplikácií a služieb vyžadovalo riadenie obchodnej logiky a spracovanie údajov, s príchodom agentov AI sa tieto funkcie môžu presunúť na intelligentné systémy, ktoré pracujú priamo s údajmi.

Preto sa v IT oddeleniach a na úrovni manažmentu čoraz častejšie diskutuje o hybridných architektúrach, v rámci ktorých sa cloudové služby dopĺňajú o AI -agentov a lokálne riešenia, čím sa znížuje závislosť od platforem SaaS.

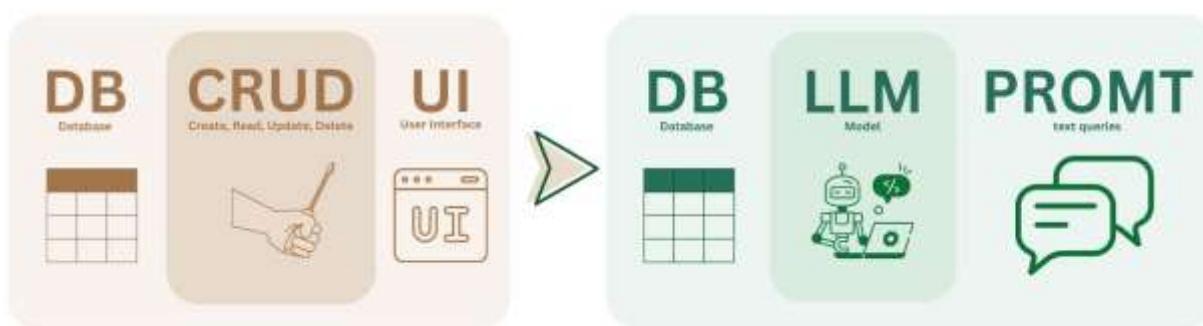
*V našom prístupe si uvedomujeme, že tradičné podnikové aplikácie alebo aplikácie SaaS sa môžu vo veku agentov výrazne zmeniť. Tieto aplikácie sú v podstate CRUD [create, read, update and delete] databázy s obchodnou logikou. V budúcnosti však túto logiku prevezmú agenti s umelou inteligenciou [46].*

- Satya Nadella, generálny riaditeľ spoločnosti Microsoft, 2024.

Prístup zameraný na údaje a využívanie agentov AI/LLM môže znížiť počet nadbytočných procesov, a

tým znížiť pracovné zaťaženie zamestnancov. Ak sú údaje správne organizované, je jednoduchšie ich analyzovať, vizualizovať a použiť pri rozhodovaní. Namiesto nekonečných reportov a kontrol získajú špecialisti prístup k aktuálnym informáciám na niekoľko kliknutí alebo pomocou agentov LLM automaticky vo forme hotových dokumentov a dashboardov.

Pri manipulácii s údajmi nám budú pomáhať nástroje umelej inteligencie (AI) a chaty LLM. V posledných rokoch sa prejavuje trend odklonu od tradičných operácií CRUD (create, read, update, delete) smerom k používaniu veľkých jazykových modelov (LLM) na správu údajov. LLM sú schopné interpretovať prirodzený jazyk a automaticky generovať vhodné databázové dotazy, čo zjednodušuje interakciu so systémami na správu údajov (obr. 2.2-3).



Obr. 2.2-3 AI nahradí a integruje úložné a databázové riešenia a postupne vytlačí tradičné aplikácie a CRUD -operácie.

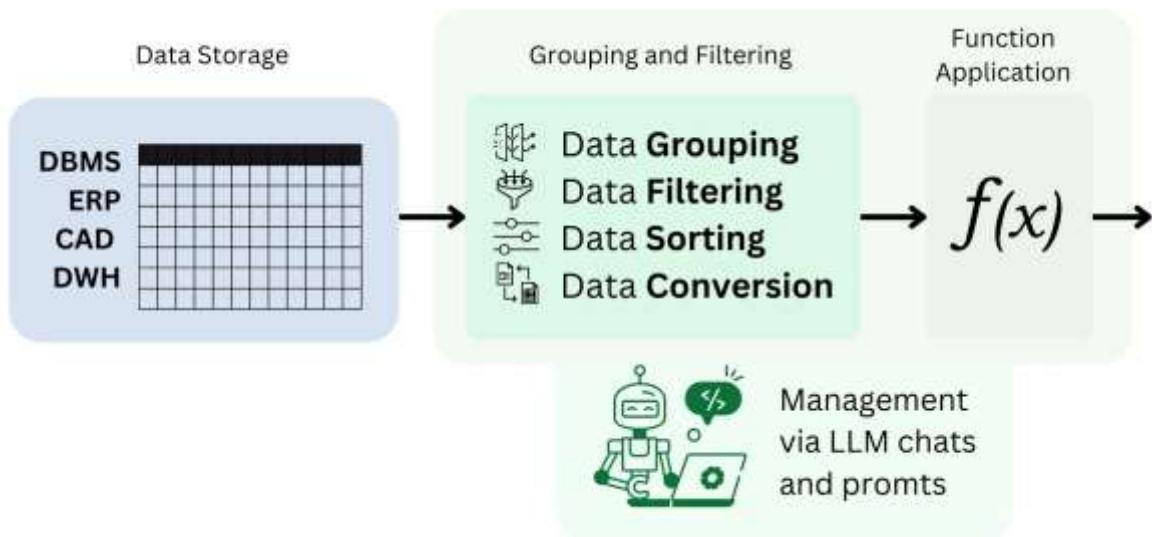
V nasledujúcich 3-6 mesiacoch bude 90% kódu písat umelá inteligencia a za 12 mesiacov by takmer všetok kód mohla generovať umelá inteligencia [47].

- Dario Amodei, generálny riaditeľ spoločnosti LLM Anthropic, marec 2025.

Napriek rýchlemu vývoju nástrojov na vývoj umelej inteligencie (napr. GitHub Copilot) budú v roku 2025 vývojári stále zohrávať klúčovú úlohu v tomto procese. Agenti AI sa stávajú čoraz užitočnejšími asistentmi: automaticky interpretujú dotazy používateľa, generujú dotazy SQL a Pandas (viac o tom v nasledujúcich kapitolách) alebo píšu kód na analýzu údajov. Umelá inteligencia tak postupne nahradza tradičné používateľské rozhrania aplikácií.

Rozširovanie modelov umelej inteligencie, ako sú jazykové modely, bude hnacou silou vývoja hybridných architektúr. Namiesto úplného upustenia od clouдовých riešení a produktov SaaS môžeme byť svedkami integrácie cloudových služieb s lokálnymi systémami správy údajov. Napríklad federatívne učenie umožňuje výkonné modely umelej inteligencie bez nutnosti presúvať citlivé údaje do cloutu. Spoločnosti si tak môžu zachovať kontrolu nad svojimi údajmi a zároveň získať prístup k

pokročilým technológiám.



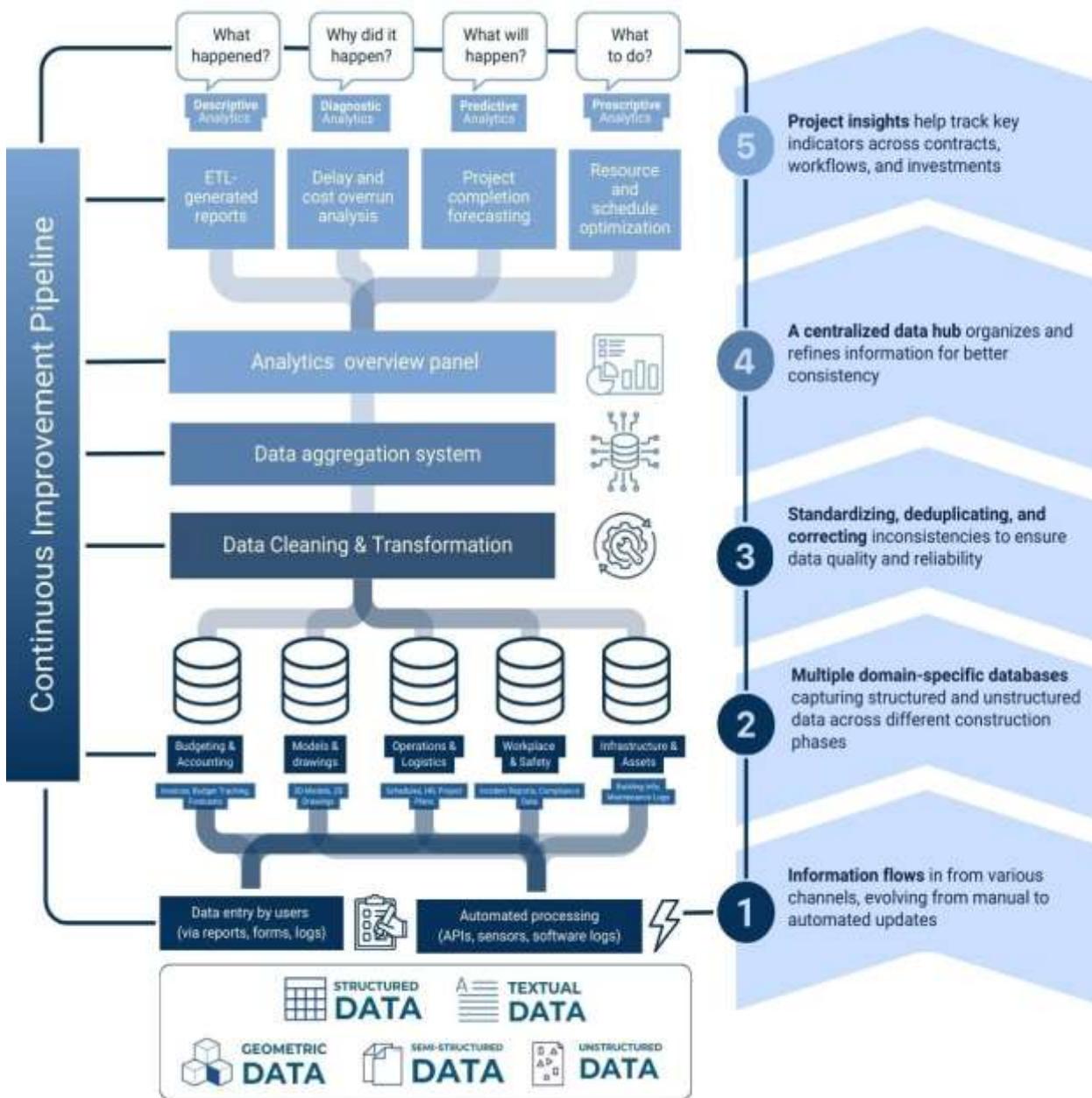
Obrázok 2.2-4 Základné operácie zoskupovania, filtrovania a triedenia, po ktorých nasleduje aplikácia funkcií, budú spracované pomocou chatu LLM.

Budúcnosť stavebníctva bude založená na kombinácii lokálnych riešení, cloudového výkonu a inteligentných modelov, ktoré budú spolupracovať na vytváraní efektívnych a bezpečných systémov správy údajov. LLM umožní používateľom bez hlbokých technických znalostí komunikovať s databázami a dátovými skladmi tak, že budú svoje dotazy formulovať v prirodzenom jazyku. Viac o agentoch LLM a AI a o tom, ako fungujú, budeme hovoriť v kapitole "Agenti LLM a štruktúrované formáty údajov".

Správne usporiadane údaje a jednoduché, ľahko použiteľné analytické nástroje s podporou LLM nielen uľahčia prácu s informáciami, ale pomôžu aj minimalizovať chyby, zvýšiť efektívnosť a automatizovať procesy.

## Od zberu údajov k rozhodovaniu: cesta k automatizácii

V ďalších častiach knihy sa budeme podrobne zaoberať tým, ako špecialisti navzájom spolupracujú a ako sa údaje stávajú základom pre rozhodovanie, automatizáciu a prevádzkovú efektívnosť. Na obrázku 2.2-5 je uvedený príklad diagramu znázorňujúceho postupnosť krokov spracovania údajov v dátovo orientovanom prístupe. Tento diagram znázorňuje potrubie neustáleho zlepšovania), ktorého časti budú podrobne rozobraté neskôr v knihe.



Obr. 2.2-5 Príklad kontinuálneho zdokonaľovania údajov: tok spracovania a analýzy údajov v stavebných projektoch.

Systém popisujúci obchodné procesy stredne veľkej spoločnosti je postavený na viacúrovňovom princípe. Zahŕňa: zber údajov, čistenie, agregáciu, analytické spracovanie a rozhodovanie na základe výsledkov. Všetky tieto etapy budeme študovať neskôr v knihe - na v teoretickom kontexte aj na praktických príkladoch:

- Na prvej úrovni prebieha **zadávanie údajov** (obr. 3.1-1). Informácie sa prijímajú manuálne (prostredníctvom hlásení, formulárov, protokolov) a v automatizovanej forme (z API, senzorov, softvérových systémov). Údaje môžu mať rôznu štruktúru: geometrickú, textovú,

neštruktúrovanú. V tejto fáze je potrebná štandardizácia, štruktúrovanie a zjednotenie informačných tokov.

- Ďalšou úrovňou je **spracovanie a transformácia údajov**. Zahŕňa procesy čistenia, odstraňovania duplicit, opravy chýb a prípravy informácií na ďalšiu analýzu (obrázok 4.2-5). Táto fáza je rozhodujúca, pretože kvalita analytických údajov priamo závisí od čistoty a presnosti údajov.
- **Údaje** potom **prúdia do špecializovaných tabuľiek, dátových rámcov alebo databáz** rozdelených podľa funkčných oblastí: rozpočet a účtovníctvo, modely a výkresy, logistika, bezpečnosť a infraštruktúra. Toto rozdelenie umožňuje jednoduchý prístup k informáciám a ich vzájomnému analýzovaniu.
- Údaje sa potom **agregujú a zobrazujú na analytickom paneli** (ukážka). Tu sa uplatňujú metódy deskriptívnej, diagnostickej, prediktívnej a preskriptívnej analýzy. To pomáha odpovedať na kľúčové otázky (obr. 1.1-4): čo sa stalo, prečo sa to stalo, čo sa stane v budúcnosti a aké opatrenia je potrebné prijať. Systém môže napríklad identifikovať oneskorenia, predpovedať dokončenie projektov alebo optimalizovať zdroje.
- Posledná úroveň vytvára **analytické závery a kľúčové ukazovatele**, ktoré pomáhajú monitorovať plnenie zmlúv, riadiť investície a zlepšovať podnikové procesy (obr. 7.4-2). Tieto informácie sa stávajú základom pre rozhodovanie a strategiu rozvoja spoločnosti.

Podobne aj údaje sa zo zberu stávajú súčasťou strategického riadenia. V nasledujúcich častiach knihy sa budeme podrobne venovať každej fáze, pričom sa zameriame na typy údajov, techniky spracovania údajov, analytické nástroje a reálne prípady využitia týchto prístupov v stavebníctve.

## Ďalšie kroky: premena chaosu na riadený systém

V tejto časti sme skúmali problémy informačných sústav a zaobrali sme sa vplyvom nadmernej zložitosti systému na výkonnosť podniku, pričom sme analyzovali prechod od štvrtej priemyselnej revolúcie k piatej, v ktorej sú kľúčové skôr údaje ako aplikácie. Videli sme, ako izolované informačné systémy vytvárajú prekážky pre zdieľanie znalostí a ako pretrvávajúca zložitosť IT prostredia znižuje produktivitu a brzdí inovácie v stavebníctve.

Na zhrnutie tejto časti stojí za to zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované prístupy vo vašich každodených úlohách:

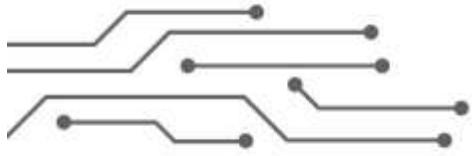
- Vizualizujte svoje informačné prostredie
  - Vytvorenie vizuálnej mapy zdrojov údajov (Miro, Figma, Canva), s ktorými pravidelne pracujete.
  - Pridajte do tejto mapy systémy a aplikácie, ktoré používate pri svojej práci.
  - Identifikácia potenciálne duplicitných funkcií a nadbytočných riešení
  - Identifikovať kritické body, kde môže dôjsť k strate alebo poškodeniu údajov počas prenosu medzi systémami.
- Zavedenie personalizovaných postupov správy údajov
  - Presun pozornosti z aplikácií na údaje ako kľúčové aktívum v procesoch

- zdokumentovanie zdrojov údajov a metodiky spracovania s cieľom zabezpečiť transparentnosť
  - Vyvinúť mechanizmy na hodnotenie a zlepšovanie kvality údajov
  - Snažte sa zabezpečiť, aby sa údaje zadávali raz a používali opakovane - to je základom efektívnej organizácie procesov
- Podporujte vo svojom tíme prístup zameraný na údaje (data-driven)
- navrhnúť používanie štandardizovaných a jednotných formátov na vzájomnú výmenu údajov
  - Pravidelne upozorňovať na problémy týkajúce sa kvality a dostupnosti údajov na stretnutiach tímu.
  - Zoznámte sa s alternatívami Open Source k nástrojom, ktoré používate na riešenie svojich problémov

Začnite v malom - vyberte si jeden konkrétny proces alebo súbor údajov, ktorý je pre vašu prácu kľúčový, a uplatnite naň prístup zameraný na údaje, pričom sa presuniete z nástrojov na údaje. Dosiahnutím úspechu v jednom pilotnom projekte získate nielen praktické skúsenosti, ale aj jasné ukážku prínosov novej metodiky pre váš tím. Pri plnení väčšiny týchto krokov, ak máte otázky, môžete požiadať o vysvetlenie a pomoc ktorúkoľvek aktuálnu LLM.

V nasledujúcich častiach knihy sa budeme podrobnejšie venovať technikám štruktúrovania a harmonizácie údajov a preskúmame praktické prístupy k integrácii heterogénnych informácií. Osobitnú pozornosť budeme venovať prechodu od rôznorodých dátových sústav k zjednoteným dátovým ekosystémom, ktoré zohrávajú kľúčovú úlohu pri digitálnej transformácii stavebníctva.





### **III ČASŤ**

## **DÁTOVÝ RÁMEC V OBCHODNÝCH PROCESOCH V STAVEBNÍCTVE**

V tretej časti sa vytvára komplexný obraz o typológií údajov v stavebníctve a metódach ich efektívnej organizácie. Analyzujú sa charakteristiky a špecifika práce so štruktúrovanými, neštruktúrovanými, pološtruktúrovanými, textovými a geometrickými údajmi v kontexte stavebných projektov. Preskúmajú sa moderné formáty ukladania a protokoly na výmenu informácií medzi rôznymi systémami používanými v odvetví. Opisujú sa praktické nástroje a techniky na prevod údajov vo viacerých formátoch do jednotného štruktúrovaného prostredia vrátane spôsobu integrácie údajov CAD (BIM). Navrhujú sa prístupy na zabezpečenie kvality údajov prostredníctvom štandardizácie a validácie, ktoré sú rozhodujúce pre presnosť stavebných výpočtov. Podrobne sa analyzujú praktické aspekty používania moderných technológií (Python Pandas, LLM -modely) s príkladmi kódu na riešenie typických problémov v stavebníctve. Zdôvodňuje sa význam vytvorenia kompetenčného centra (CoE) ako organizačnej štruktúry na koordináciu a štandardizáciu prístupov k riadeniu informácií.

## KAPITOLA 3.1.

# DÁTOVÉ TYPY V STAVEBNÍCTVE

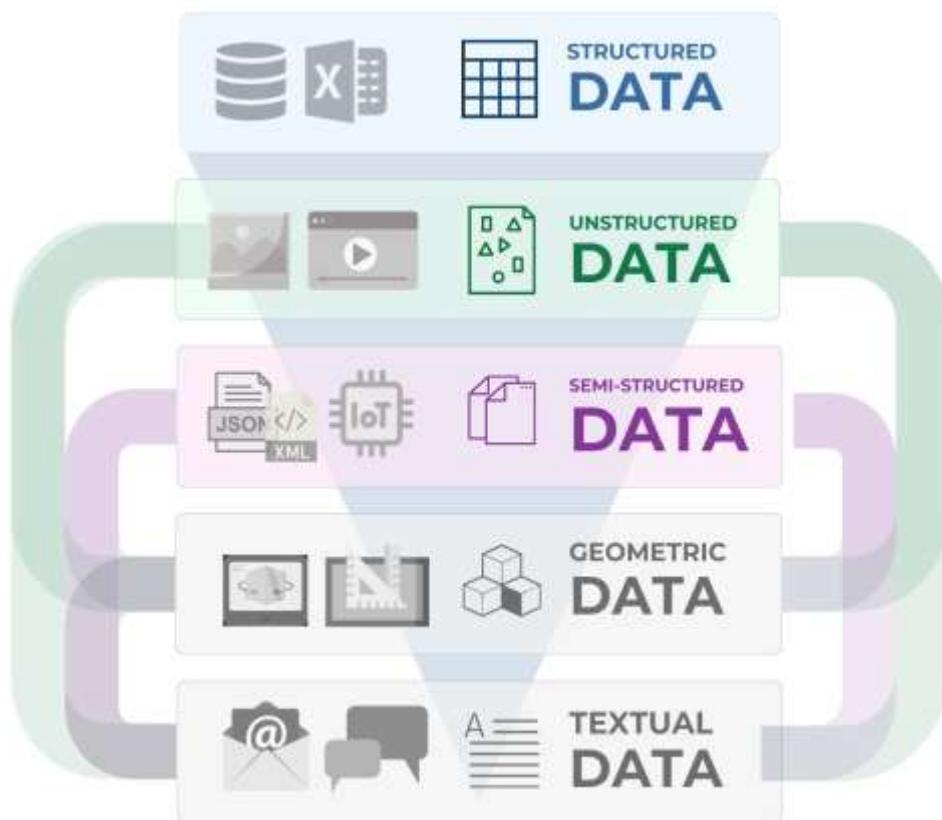
### Najdôležitejšie typy údajov v stavebníctve

V modernom stavebníctve sa systémy, aplikácie a dátové sklady spoločností aktívne napĺňajú informáciami a údajmi rôznych typov a formátov (obr. 3.1-1). Pozrite sa bližšie na hlavné typy údajov, ktoré tvoria informačné prostredie modernej spoločnosti pôsobiacej v stavebníctve:

- **Štruktúrované údaje:** tieto údaje majú jasnú organizačnú štruktúru, napr. tabuľky Excel a relačné databázy.
- **Neštruktúrované údaje:** ide o informácie, ktoré nie sú usporiadane podľa prísnych pravidiel. Príkladmi takýchto údajov sú text, video, fotografie a zvukové záznamy.
- **Voľne štruktúrované údaje:** tieto údaje zaujímajú medzipolohu medzi štruktúrovanými a neštruktúrovanými údajmi. Obsahujú prvky štruktúry, ale táto štruktúra nie je vždy jasná alebo je často opísaná prostredníctvom rôznych schém. Príkladmi pološtruktúrovaných údajov v stavebníctve sú: technické špecifikácie, projektová dokumentácia alebo správy o priebehu prác.
- **Textové údaje:** zahŕňajú všetko, čo pochádza z ústnej a písomnej komunikácie, napríklad e-maily, prepisy stretnutí a schôdzok.
- **Geometrické údaje:** tieto údaje pochádzajú z programov CAD, v ktorých špecialisti vytvárajú geometrické údaje prvkov projektu na vizualizáciu, potvrdenie objemových hodnôt alebo kontrolu kolízii.

Je dôležité poznamenať, že geometrické a textové (alfanumerické) údaje nie sú samostatnou kategóriou, ale môžu sa vyskytovať vo všetkých troch typoch údajov. Geometrické údaje, napríklad, môžu byť súčasťou štruktúrovaných údajov (parametrické formáty CAD) aj neštruktúrovaných údajov (naskenované výkresy). Textové údaje môžu byť podobne organizované v databázach (štruktúrované údaje) aj existovať ako dokumenty bez jasnej štruktúry.

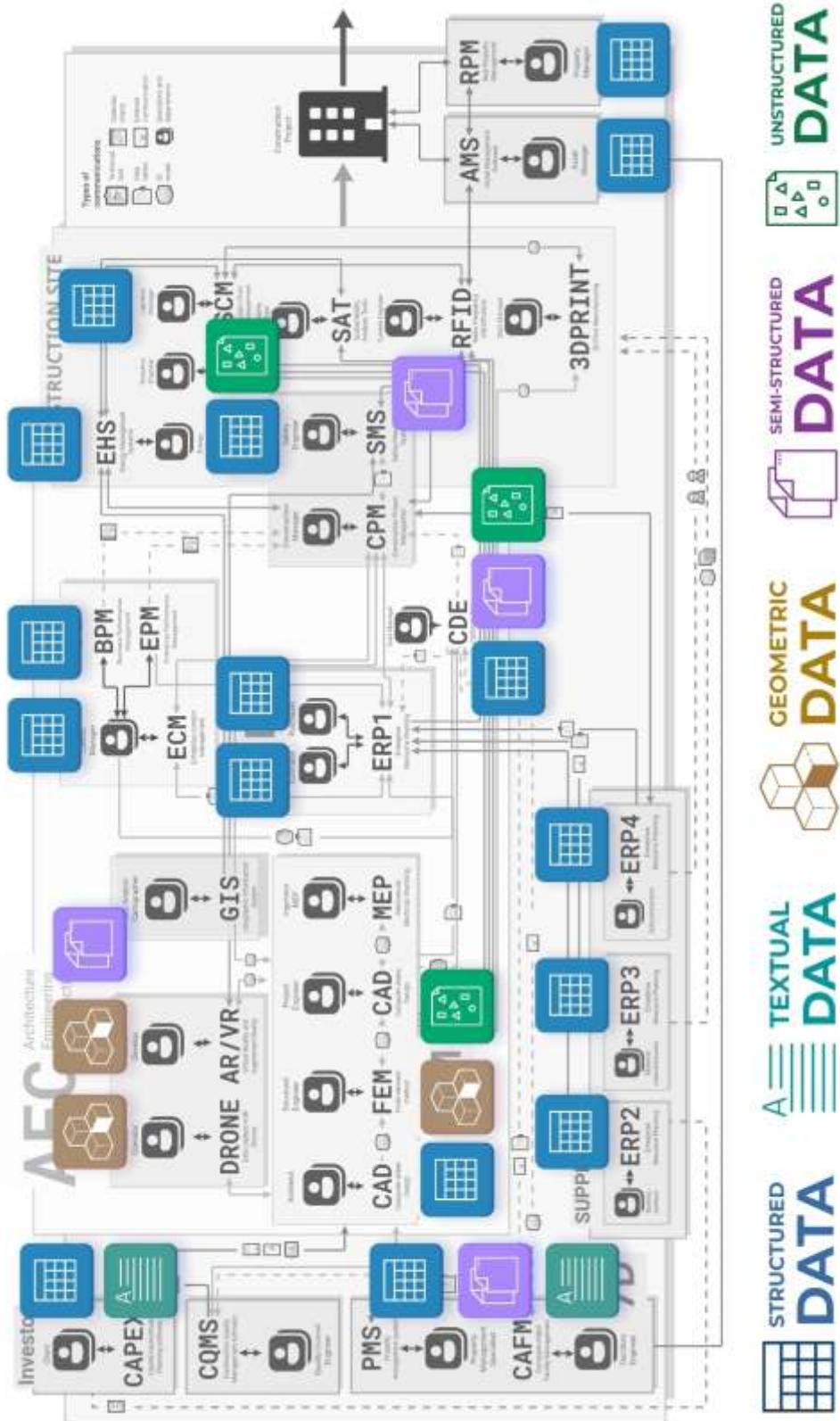
Každý typ údajov v stavebnej spoločnosti je jedinečným prvkom v mozaike informačných aktív spoločnosti. Od neštruktúrovaných údajov, ako sú obrázky zo staveniska a zvukové záznamy z porád, až po štruktúrované záznamy vrátane tabuľiek a databáz, každý prvek zohráva dôležitú úlohu pri formovaní informačného prostredia spoločnosti.



Obr. 3.1-1 Inžinieri a správcovia údajov sa musia naučiť pracovať so všetkými typmi údajov používaných v stavebníctve.

Tu je vzorový zoznam len niektorých systémov a súvisiacich typov údajov (obrázok 3.1-2), ktoré sa používajú v stavebníctve:

- **ERP** (Enterprise Resource Planning) - spracúva všeobecne štruktúrované údaje, ktoré pomáhajú spravovať podnikové zdroje a integrovať rôzne podnikové procesy.
- **CAD** (Computer-Aided Design) v kombinácii s **BIM** (Building Information Modeling) - využíva geometrické a pološtruktúrované údaje na navrhovanie a modelovanie stavebných projektov, čím zabezpečuje presnosť a konzistentnosť informácií vo fáze návrhu.
- **GIS** (Geografické informačné systémy) - pracuje s geometrickými a štruktúrovanými údajmi s cieľom vytvárať a analyzovať kartografické údaje a priestorové vzťahy.
- **RFID** (Radio-Frequency Identification) - využíva pološtruktúrované údaje na efektívne sledovanie materiálov a zariadení na stavenisku pomocou rádiovfrekvenčnej identifikácie.
- **ECM** (Engineering Content Management) je systém na správu technických údajov a dokumentácie vrátane pološtruktúrovaných a neštruktúrovaných údajov, ako sú technické výkresy a konštrukčné dokumenty.

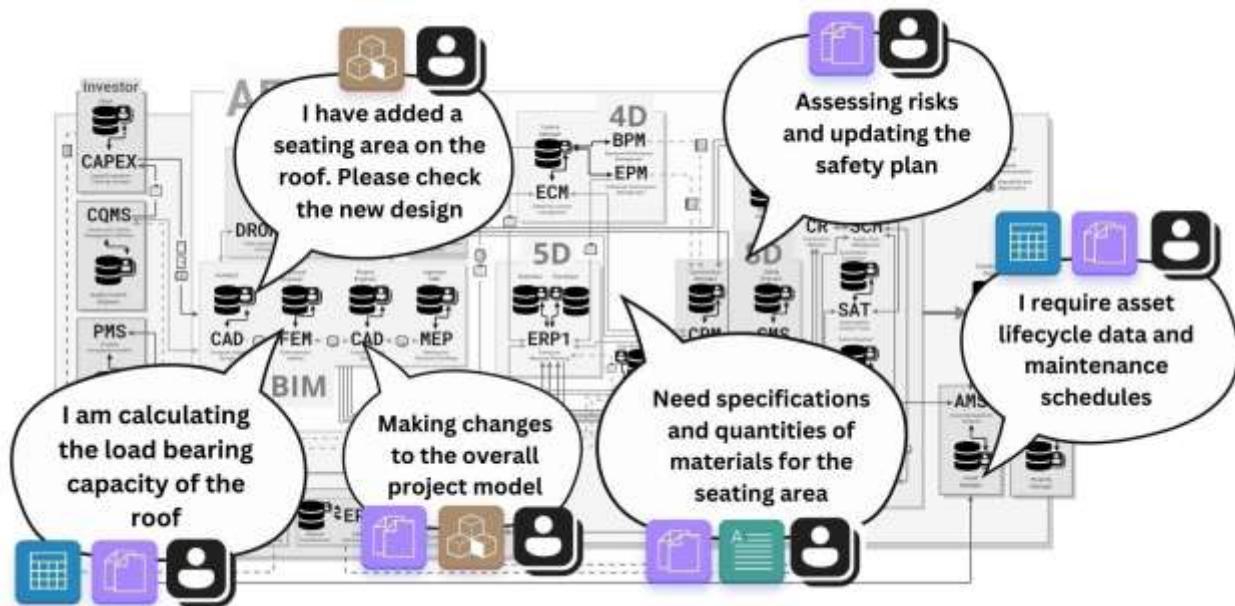


Obr. 3.1-2 Rôzne formáty a údaje napĺňajú rôzne systémy, čo si vyžaduje preklad do podoby vhodnej na komplexnú integráciu.

Tieto a mnohé ďalšie systémy spoločnosti spravujú širokú škálu údajov, od štruktúrovaných tabuľkových údajov až po komplexné geometrické modely, a poskytujú integrovanú spoluprácu v procesoch navrhovania, plánovania a riadenia výstavby.

V príklade zjednodušeného dialógu (obr. 3.1-3) si špecialisti stavebného projektu vymieňajú rôzne typy údajov:

- ⦿ **Architekt:** "Vzhľadom na želanie klienta som na strechu pridal priestor na sedenie. Prosím, pozrite si nový návrh" (geometrické údaje - model).
- ⦿ **Stavebný inžinier:** "Projekt bol prijatý. Vypočítavam nosnosť strechy pre novú rekreačnú oblast" (štruktúrované a pološtruktúrované údaje - výpočtové tabuľky).
- ⦿ **Manažér nákupu:** "Potrebujete špecifikácie a množstvá materiálov pre rekreačnú oblasť, aby ste mohli zorganizovať nákup" (textové a pološtruktúrované údaje - zoznamy a špecifikácie).
- ⦿ **Inžinier pre bezpečnosť a ochranu zdravia:** "Dostal som údaje o novej oblasti. Posudzujem riziká a aktualizujem bezpečnostný plán" (pološtruktúrované údaje - dokumenty a plány).
- ⦿ **Špecialista na BIM - modelovanie:** "Vykonávanie zmien v celkovom modeli projektu s cieľom upraviť pracovnú dokumentáciu" (geometrické údaje a pološtruktúrované údaje).
- ⦿ **Projektový manažér:** "Novú oddychovú zónu začleňujem do harmonogramu prác. Aktualizujem harmonogramy a zdroje v systéme riadenia projektu" (štruktúrované a pološtruktúrované údaje - harmonogramy a plány).
- ⦿ **Špecialista na údržbu zariadení (FM):** "Pripravujem údaje pre budúcu údržbu rekreačnej oblasti a vkladám ich do systému správy majetku" (štruktúrované a pološtruktúrované údaje - pokyny a plány údržby).



Obr. 3.1-3 Komunikácia medzi špecialistami prebieha na textovej aj dátovej úrovni.

Každý odborník pracuje s rôznymi typmi údajov, aby zabezpečil efektívnu tímovú spoluprácu a úspech projektu. Pochopenie rozdielov medzi štruktúrovanými, pološtruktúrovanými a neštruktúrovanými údajmi vám umožní rozpoznať jedinečnú úlohu, ktorú každý typ zohráva v digitálnych obchodných procesoch. Je dôležité nielen vedieť, že existujú rôzne formy údajov, ale aj pochopiť, ako, kde a prečo

sa používajú.

Nie je to tak dávno, čo sa myšlienka spojenia takýchto rôznorodých údajov zdala byť ambiciozna, ale ľažko realizovateľná. Dnes je už súčasťou každodennej praxe. Integrácia údajov rôznych schém a štruktúr sa stala neoddeliteľnou súčasťou architektúry moderných informačných systémov.

V nasledujúcich kapitolách sa budeme podrobne zaoberať kľúčovými normami a prístupmi, ktoré umožňujú spojiť štruktúrované, pološtruktúrované a neštruktúrované údaje do jedného uceleného prehľadu. Osobitnú pozornosť budeme venovať štruktúrovaným údajom a relačným databázam ako hlavným mechanizmom na ukladanie, spracovanie a analýzu informácií v stavebníctve.

## Štruktúrované údaje

V stavebníctve pochádzajú informácie z mnohých zdrojov - výkresov, špecifikácií, rozpisov a správ. Aby bolo možné tento tok informácií efektívne riadiť, je potrebné ich štruktúrovať. Štruktúrované údaje umožňujú usporiadať informácie v pohodnej, čitateľnej a prístupnej forme.

Podľa 5. výročnej správy o stavebných technológiach spoločnosti JB Knowledge [17] 67% odborníkov na riadenie stavebných projektov sleduje a hodnotí výkonnosť práce manuálne alebo pomocou tabuliek.

Medzi najbežnejšie formáty štruktúrovaných údajov patria XLSX a CSV. Sú široko používané na ukladanie, spracovanie a analýzu informácií v tabuľkách. V takýchto tabuľkách sú údaje prezentované vo forme riadkov a stĺpcov, čo uľahčuje ich čítanie, úpravu a analýzu.

**XLSX**, formát vytvorený spoločnosťou Microsoft, je založený na použití štruktúr XML a archivuje sa pomocou algoritmu ZIP. Hlavné vlastnosti tohto formátu:

- Podpora zložitých vzorcov, grafov a makier.
- Možnosť ukladať údaje do rôznych hárkov, ako aj formátovať informácie.
- Optimalizované pre Microsoft Excel, ale kompatibilné s inými kancelárskymi balíkmi.

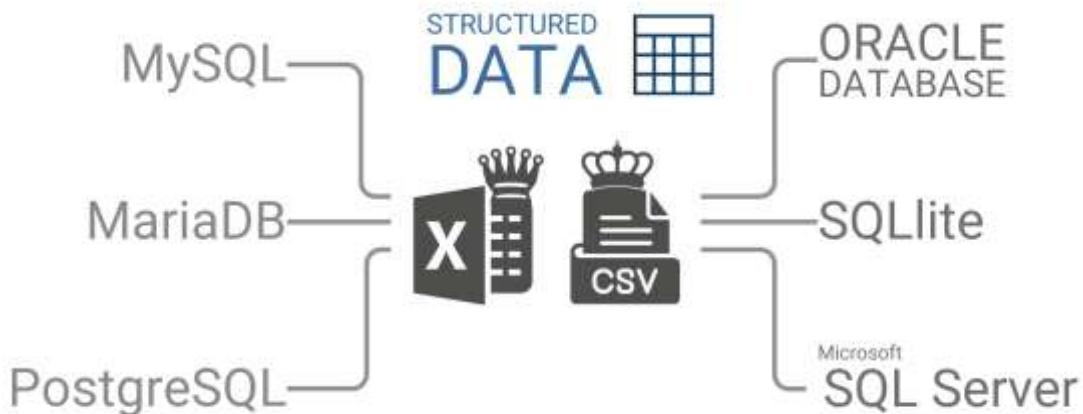
**Formát CSV** je obyčajný textový súbor, v ktorom sú hodnoty oddelené čiarkami, stredníkmi alebo inými oddelovacími znakmi. Hlavné výhody:

- Univerzálna kompatibilita s rôznymi programami a operačnými systémami.
- Jednoduchý import/export do databáz a analytických systémov.
- Jednoduché spracovanie aj v textových editoroch.

CSV však nepodporuje vzorce a formátovanie, takže jeho hlavným využitím je výmena údajov medzi systémami a hromadná aktualizácia informácií. Vďaka svojej univerzálnosti a nezávislosti od platformy sa CSV stal obľúbeným nástrojom na prenos údajov v heterogénnych IT prostrediach.

Dva formáty XLSX a CSV slúžia ako prepojenie medzi rôznymi systémami pracujúcimi so štruktúrovanými údajmi (obrázok 3.1-4). Sú užitočné najmä pri úlohách, kde je dôležitá čitateľnosť,

ručná úprava a základná kompatibilita



Obr. 3.1-4 Formáty XLSX a CSV sú prepojení medzi rôznymi systémami, ktoré pracujú so štruktúrovanými údajmi.

Nezávislosť na platforme robí z CSV najobľúbenejší formát na prenos údajov v heterogénnych IT prostrediach a systémoch.

XLSX a CSV však nie sú určené na vysokovýkonné výpočty alebo dlhodobé ukladanie veľkého množstva údajov. Na takéto účely sa používajú modernejšie štruktúrované formáty, ako napríklad Apache Parquet, Apache ORC, Feather, HDF5. Týmto formátom sa budeme podrobnejšie venovať v kapitole "Ukladanie veľkých objemov údajov: Analýza populárnych formátov a ich efektívnosť" v 9. časti tejto knihy.

V praxi sa Excel s formátom XLSX častejšie používa na drobné úlohy a automatizáciu rutinných procesov. Zložitejšie scenáre si vyžadujú použitie systémov na správu údajov, ako sú ERP, PMIS CAFM, CPM, SCM a iné (obr. 3.2-1). Tieto systémy uchovávajú štruktúrované údaje, na ktorých je založená organizácia a riadenie informačných tokov spoločnosti.

Moderné informačné systémy na správu údajov používané v stavebníctve sa spoliehajú na štruktúrované údaje usporiadane vo forme tabuľiek. Na účely spoľahlivej, škálovateľnej a holistickej správy veľkých objemov informácií sa vývojári aplikácií a systémov obracajú na relačné systémy riadenia databáz (RDBMS).

## Relačné databázy RDBMS a dotazovací jazyk SQL

**Relačné databázy (RDBMS)** sú systémy na ukladanie údajov, ktoré organizujú informácie do tabuľiek s definovanými vzťahmi medzi nimi.

Údaje usporiadane v databázach (RDBMS) nie sú len digitálnymi informáciami, ale aj základom transakcií a interakcií medzi rôznymi systémami.

Tu je niekoľko najbežnejších systémov riadenia relačných databáz (RDBMS) (obrázok 3.1-5):

- **MySQL** (Open Source) je jeden z najpopulárnejších RDBMS, ktorý je súčasťou LAMP stacku (Linux, Apache, MySQL, PHP /Perl/Python). Vďaka svojej jednoduchosti a vysokému výkonu sa široko používa pri vývoji webových stránok.
- **PostgreSQL** (Open Source) je výkonný objektovo-relačný systém známy svojou spoľahlivosťou a pokročilými funkiami. Je vhodný pre komplexné podnikové riešenia.
- **Server Microsoft SQL** je komerčný systém od spoločnosti Microsoft, ktorý sa široko používa v podnikových prostrediach vďaka svojej integrácii s inými produktmi spoločnosti a vysokej úrovni zabezpečenia.
- **Databáza Oracle** je jedným z najvýkonnejších a najspoľahlivejších DBMS používaných vo veľkých podnikoch a kritických aplikáciách.
- **IBM DB2** - určený pre veľké spoločnosti, poskytuje vysoký výkon a odolnosť voči chybám.
- **SQLite** (Open Source) je ľahká vstavaná databáza, ideálna pre mobilné aplikácie a samostatné systémy, ako je napríklad návrhový softvér CAD (BIM).

Oblúbené systémy správy databáz v stavebníctve - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 a SQLite - pracujú so štruktúrovanými údajmi. Všetky tieto DBMS predstavujú výkonné a flexibilné riešenia na správu širokého spektra podnikových procesov a aplikácií, od malých webových stránok až po rozsiahle podnikové systémy (obr. 3.2-1).

Podľa portálu Statista [48] budú relačné systémy riadenia databáz (RDBMS) v roku 2022 predstavovať približne 72% všetkých používaných DBMS.

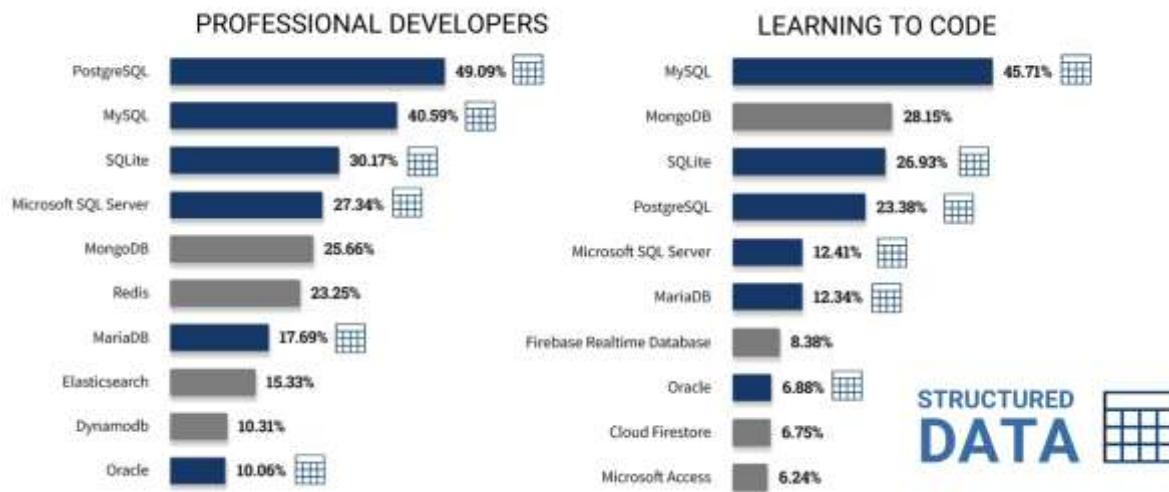
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

Obr. 3.1-5 Oblúbenosť používania štruktúrovaných databáz (vyznačené modrou farbou) v rebríčku DBMS (podľa [49]).

Inštalácia databáz s otvoreným zdrojovým kódom je pomerne jednoduchá aj bez rozsiahlych technických znalostí. Systémy s otvoreným zdrojovým kódom, ako napríklad PostgreSQL, MySQL alebo SQLite, sú k dispozícii zadarmo a fungujú na väčšine operačných systémov: Windows, MacOS a Linux. Stačí navštíviť oficiálnu webovú stránku projektu, stiahnuť si inšalačný program a postupovať podľa pokynov. Vo väčšine prípadov inštalácia netrvá dlhšie ako 10 až 15 minút. Jednu takúto databázu si vymodelujeme a vytvoríme vo štvrtej časti knihy (obr. 4.3-8).

Ak vaša spoločnosť využíva clouдовé služby (napríklad Amazon Web Services, Google Cloud alebo Microsoft Azure), môžete databázu nasadiť na niekoľko kliknutí - platforma vám ponúkne pripravené šablóny na inštaláciu. Vďaka otvorenosti kódu sa takéto databázy dajú ľahko prispôsobiť vašim úlohám a obrovská komunita používateľov vám vždy pomôže nájsť riešenie akéhokoľvek problému.

RDBMS zostávajú základom pre množstvo podnikových aplikácií a analytických platforiem (obrázok 3.1-6), ktoré umožňujú spoločnostiam efektívne ukladať, spracovávať a analyzovať údaje, a teda prijímať informované a včasné rozhodnutia.



Obr. 3.1-6 Prieskum vývojárov na StackOverFlow (najväčšie IT fórum) o tom, ktoré databázy používali v minulom roku a ktoré chcú používať v budúcom roku (RDBMS sú zvýraznené modrou farbou) (podľa [50]).

Systémy RDBMS poskytujú spoločnosť, konzistenciu údajov, podporu transakcií a používajú výkonný dopytovací jazyk SQL (Structured Query Language), ktorý sa často používa v analytike a umožňuje jednoducho získať, upravovať a analyzovať informácie uložené v databázach. Jazyk SQL je hlavným nástrojom na prácu s údajmi v relačných systémoch.

## SQL - dopyty v databázach a nové trendy

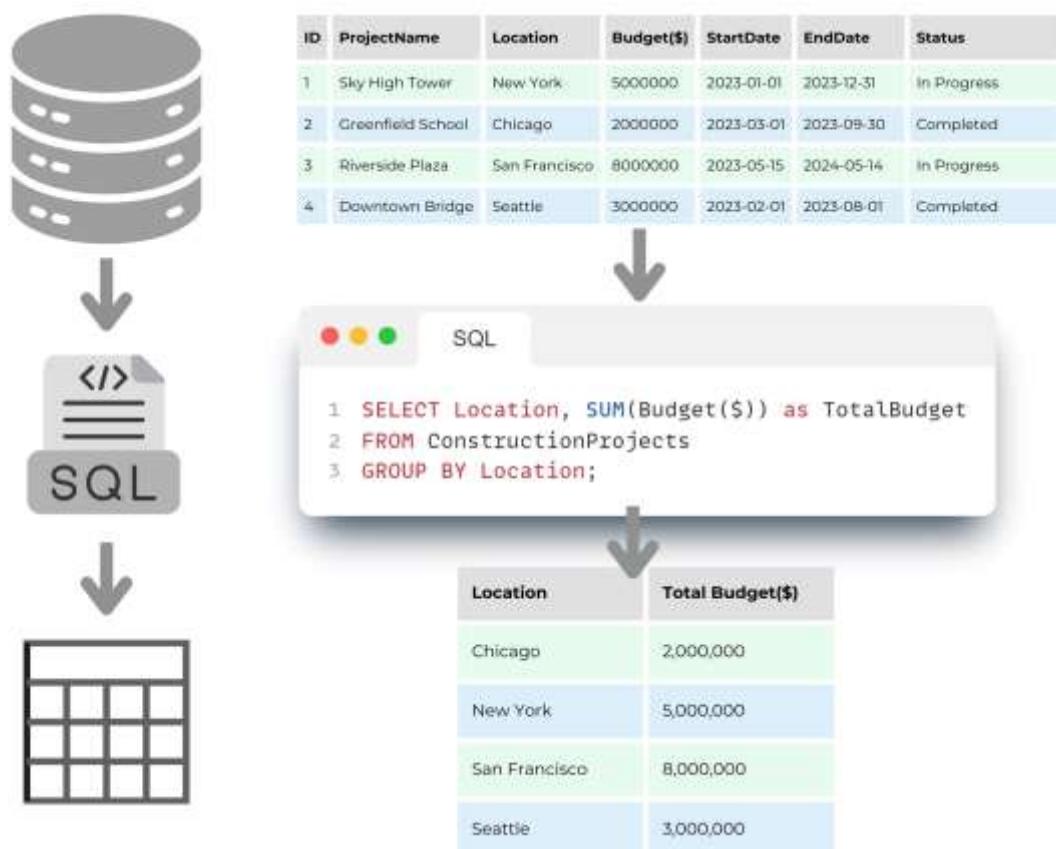
Hlavnou výhodou jazyka SQL, ktorý sa často používa v relačných databázach, oproti iným typom správy informácií (napríklad pomocou klasických tabuľiek Excelu) je podpora veľmi veľkých objemov databáz pri vysokej rýchlosťi spracovania dotazov.

Štruktúrovaný dotazovací jazyk (SQL) je špecializovaný programovací jazyk určený na ukladanie, spracovanie a analýzu informácií v relačných databázach. Jazyk SQL sa používa na vytváranie, správu a prístup k údajom, pričom umožňuje efektívne vyhľadávať, filtrovať, kombinovať a agregovať informácie. Slúži ako kľúčový nástroj na prístup k údajom a poskytuje pohodlný a formalizovaný spôsob interakcie s informačnými skladmi.

Vývoj systémov SEQUEL-SQL prechádza cez významné produkty a spoločnosti, ako sú Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL a MySQL, a vrcholí vznikom SQLite a MariaDB [51]. SQL poskytuje tabuľkové možnosti, ktoré sa nenachádzajú v programe Excel, vďaka čomu je manipulácia s údajmi škálovateľnejšia a ľahko automatizovateľná:

- **Vytváranie a správa dátových štruktúr (DDL):** V SQL môžete vytvárať, upravovať a odstraňovať tabuľky v databáze, vytvárať medzi nimi prepojenia a definovať štruktúry ukladania údajov. Excel naproti tomu pracuje s pevne stanovenými hárkami a bunkami bez jasne definovaných vzťahov medzi hárkami a dátovými súbormi.

- **Manipulácia s údajmi (DML):** SQL umožňuje hromadne pridávať, upravovať, odstraňovať a vyhľadávať údaje vysokou rýchlosťou, vykonávať zložité dotazy s filtrovaním, triedením a spájaním tabuľiek (obrázok 3.1-7). V programe Excel si spracovanie veľkého množstva informácií vyžaduje manuálne úkony alebo špeciálne makrá, čo spomaľuje proces a zvyšuje pravdepodobnosť chýb.
- **Riadenie prístupu (DCL):** SQL umožňuje rozlísiť prístupové práva k údajom pre rôznych používateľov a obmedziť možnosť upravovať alebo zobrazovať informácie. Na druhej strane v programe Excel je prístup buď zdieľaný (pri prenose súboru), alebo si vyžaduje zložité nastavenia so zdieľaním oprávnení prostredníctvom cloudových služieb.



Obr. 3.1-7 Príklad DML v SQL: rýchle spracovanie, zoskupovanie a agregácia pomocou niekoľkých riadkov kódu na automatické spracovanie údajov.

Excel uľahčuje prácu s údajmi vďaka svojej vizuálnej a intuitívnej štruktúre. S rastúcim množstvom údajov však výkonnosť programu Excel klesá. Excel tiež naráža na limity množstva údajov, ktoré môže ukladať - maximálne jeden milión riadkov - a výkon sa znižuje dlho pred dosiahnutím tohto limitu. Takže zatiaľ čo Excel sa javí ako vhodnejší na vizualizáciu a manipuláciu s malým množstvom údajov, SQL je vhodnejší na prácu s veľkými súbormi údajov.

Ďalšou etapou vo vývoji štruktúrovaných údajov bol vznik stĺpcových databáz (Columnar Databases), ktoré sú alternatívou k tradičným relačným databázam, najmä pokial' ide o výrazne väčšie objemy údajov a analytické výpočty. Na rozdiel od riadkových databáz, kde sa údaje ukladajú po riadkoch,

stípcové databázy zaznamenávajú informácie po stípcach. V porovnaní s klasickými databázami to umožňuje:

- Zníženie úložného priestoru efektívou kompresiou jednotných údajov v stípcach.
- Zrýchlenie analytických dotazov, pretože sa načítajú len požadované stípce, nie celá tabuľka.
- Optimalizácia veľkých dát a dátových skladov, napr. architektúra dátového skladu.

O stípcových databázach, Pandas DataFrame, Apache Parquet, HDF5, ako aj o vytváraní na nich založených veľkokapacitných dátových -úložisk na účely analýzy a spracovania údajov si povieme viac v nasledujúcich kapitolách tejto knihy - "DataFrame: univerzálny formát tabuľkových údajov" a "Formáty ukladania údajov a práca s Apache Parquet: DWH - dátové skladby a architektúra Data Lakehouse".

## Neštruktúrované údaje

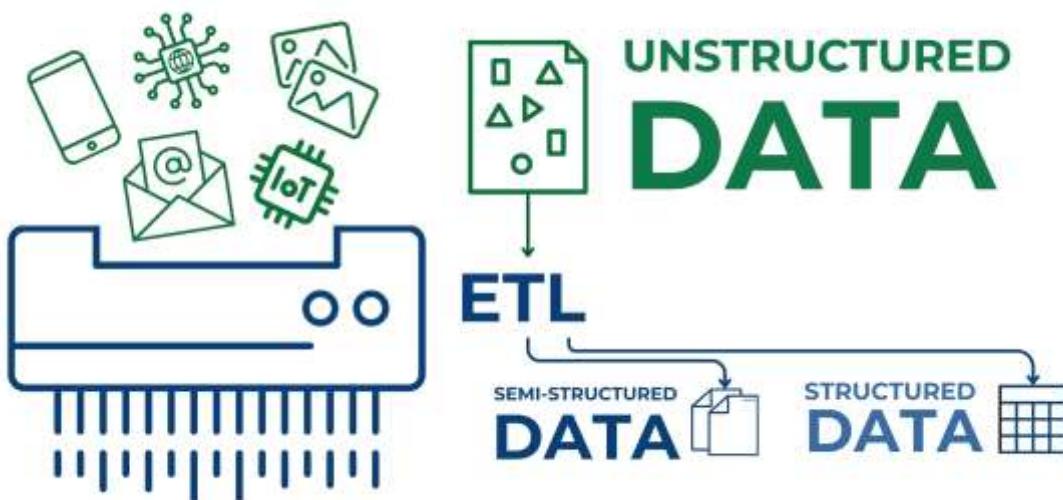
Hoci väčšina údajov používaných v aplikáciách a informačných systémoch má štruktúrovanú podobu, väčšina informácií generovaných v stavebníctve má podobu neštruktúrovaných údajov - obrázkov, videí, textových dokumentov, zvukových záznamov a iných foriem obsahu. Platí to najmä vo fáze výstavby, prevádzky a technického dozoru, kde prevládajú vizuálne a textové informácie.

**Neštruktúrované údaje** sú informácie, ktoré nemajú vopred definovaný model alebo štruktúru, nie sú usporiadané do tradičných riadkov a stípcov ako v databázach alebo tabuľkách.

Vo všeobecnosti možno neštruktúrované údaje rozdeliť do dvoch kategórií:

- Neštruktúrované údaje vytvorené človekom, ktoré zahŕňajú rôzne typy obsahu vytvoreného človekom: textové dokumenty, e-maily, obrázky, videá atď.
- Strojovo generované neštruktúrované údaje sú vytvárané zariadeniami a snímačmi: ide napríklad o súbory denníkov, údaje GPS, výsledky internetu vecí (IoT) a iné telemetrické informácie zo staveniska.

Na rozdiel od štruktúrovaných údajov, ktoré sú pohodlne usporiadané v tabuľkách a databázach, neštruktúrované údaje si vyžadujú ďalšie kroky spracovania pred ich integráciou do informačných systémov (obr. 3.1-8). Využívanie technológií na automatizovaný zber, analýzu a transformáciu takýchto údajov otvára nové možnosti na zvýšenie efektívnosti výstavby, zníženie počtu chýb a minimalizáciu vplyvu ľudského faktora.



Obrázok 3.1-8 Spracovanie neštruktúrovaných údajov začína ich transformáciou na pološtruktúrované a štruktúrované údaje.

Neštruktúrované údaje tvoria až 80% všetkých informácií [52], s ktorými sa odborníci vo firmách stretávajú, preto sa ich typom a spracovaním budeme podrobne zaoberať na príkladoch v nasledujúcich kapitolách knihy.

Na uľahčenie diskusie sú textové údaje rozdelené do samostatných kategórií. Hoci ide o typ skôr neštruktúrovaných údajov, ich význam a rozšírenosť v stavebníctve si vyžadujú osobitnú pozornosť.

### Textové údaje: medzi neštruktúrovaným chaosom a štruktúrovaným chaosom

Textové údaje v stavebníctve zahŕňajú širokú škálu formátov a typov informácií, od papierových dokumentov až po neformálne spôsoby komunikácie, ako sú listy, rozhovory, pracovná korešpondencia a ústne stretnutia na stavenisku. Všetky tieto textové údaje sú nositeľmi dôležitých informácií pre riadenie stavebných projektov, od podrobnosti o projektových rozhodnutiach a zmenách v plánoch až po diskusie o bezpečnostných otázkach a rokovania s dodávateľmi a zákazníkmi (obrázok 3.1-9).



Obr. 3.1-9 Textové údaje, jeden z najobľúbenejších typov informácií používaných pri komunikácii medzi účastníkmi projektu.

Textové informácie môžu byť formalizované aj neštruktúrované. Formalizované údaje zahŕňajú dokumenty vo formáte Word (.doc,.docx), PDF, ako aj textové súbory zápisníc zo zasadnutí (.txt). Neformalizované údaje zahŕňajú korešpondenciu prostredníctvom messengerov a e-mailov, prepisy stretnutí (Teams, Zoom, Google Meet) a zvukové záznamy diskusií (.mp3,.wav), ktoré si vyžadujú konverziu na text.

Zatiaľ čo písomné dokumenty, ako sú oficiálne žiadosti, zmluvné podmienky a e-maily, už zvyčajne majú určitú štruktúru, ústna komunikácia a pracovná korešpondencia často zostávajú neštruktúrované, čo stáhuje ich analýzu a integráciu do systémov riadenia projektov.

Kľúčom k efektívnej správe textových údajov je ich prevod do štruktúrovaného formátu. To umožňuje automatickú integráciu spracovaných informácií do existujúcich systémov, ktoré už pracujú so štruktúrovanými údajmi.



Obr. 3.1-10 Konverzia textového obsahu na štruktúrované údaje.

Aby bolo možné efektívne využívať textové informácie, musia byť automaticky prevedené do štruktúrovanej podoby (obrázok 3.1-10). Tento proces zvyčajne zahrňa niekoľko krokov:

- **Rozpoznávanie textu (OCR)** - prevod obrázkov dokumentov a výkresov do strojovo čitateľného formátu.
- **Analýza textu (NLP)** - automatická identifikácia kľúčových parametrov (dátumy, sumy a čísla relevantné pre projekt).
- **Klasifikácia údajov** - kategorizácia informácií (finančie, logistika, riadenie rizík).

Po rozpoznaní a klasifikácii možno už štruktúrované údaje integrovať do databáz a používať ich v automatizovaných systémoch výkazníctva a riadenia.

## Čiastočne štruktúrované a voľne štruktúrované údaje

Čiastočne štruktúrované údaje obsahujú určitú úroveň organizácie, ale nemajú prísnu schému alebo štruktúru. Hoci takéto informácie obsahujú štruktúrované prvky (napr. dátumy, mená zamestnancov a zoznamy splnených úloh), formát prezentácie sa môže značne lísiť od projektu k projektu alebo dokonca od jedného zamestnanca k druhému. Príkladom takýchto údajov sú časové denníky, správy o priebehu prácu a harmonogramy, ktoré môžu byť prezentované v rôznych formátoch.

Pološtruktúrované údaje sa analyzujú ľahšie ako neštruktúrované údaje, ale vyžadujú si ďalšie spracovanie na integráciu do štandardizovaných systémov riadenia projektov.

Práca s pološtruktúrovanými údajmi, ktoré sa vyznačujú neustále sa meniacou štruktúrou, predstavuje značné výzvy. Variabilita štruktúry údajov si totiž vyžaduje samostatné individuálne prístupy k spracovaniu a analýze každého zdroja pološtruktúrovaných údajov.

Zatiaľ čo práca s neštruktúrovanými údajmi si vyžaduje veľa úsilia, spracovanie pološtruktúrovaných údajov možno vykonať pomocou relatívne jednoduchých metód a nástrojov.

**Slabo štruktúrované** údaje je všeobecnejší pojem, ktorý opisuje údaje s minimálnou alebo neúplnou štruktúrou. Najčastejšie ide o textové dokumenty, chaty, e-mails, v ktorých sa nachádzajú niektoré metadáta (napr. dátum, odosielateľ), ale väčšina informácií je prezentovaná chaoticky.

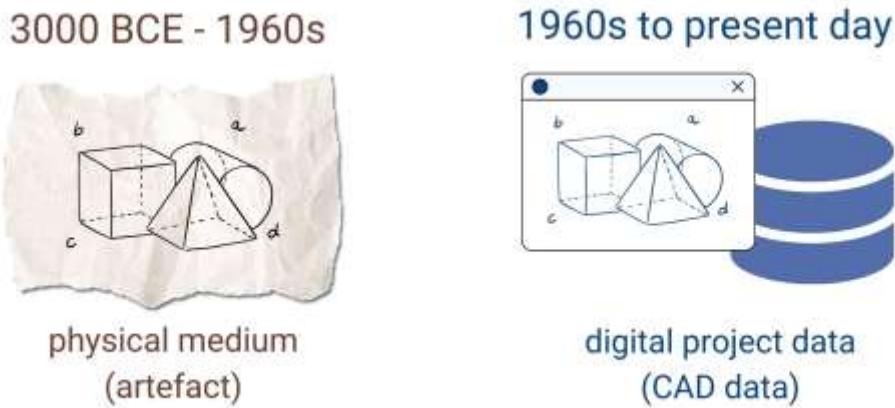
Vo výstavbe sa voľne štruktúrované údaje vyskytujú v rôznych procesoch. Môžu zahŕňať napríklad:

- Odhady a cenové ponuky - tabuľky s údajmi o materiáloch, množstvách a nákladoch, ale bez jednotného formátu.
- Výkresy a technické schémy - súbory vo formáte PDF alebo DWG, ktoré obsahujú textové poznámky a metadáta, ale nemajú striktne stanovenú štruktúru.
- Pracovné harmonogramy - údaje z MS Project, Primavera P6 alebo iných systémov, ktoré môžu mať inú štruktúru exportu.
- CAD (BIM -modely) - obsahujú prvky konštrukcie, ale reprezentácia údajov závisí od softvéru a normy projektu.

Geometrické údaje, vytvorené systémami CAD, možno kategorizovať rovnakým spôsobom ako pološtruktúrované údaje. Geometrické údaje CAD (BIM) však budeme klasifikovať ako samostatný typ údajov, pretože podobne ako s textovými údajmi s nimi možno v podnikových procesoch často zaobchádzať ako so samostatným typom údajov.

## Geometrické údaje a ich použitie

Zatiaľ čo metadáta o prvkoch projektu sú takmer vždy uložené vo forme tabuľiek, štruktúrovaných alebo voľne štruktúrovaných formátov, geometrické údaje o prvkoch projektu v sa vo väčšine prípadov vytvárajú pomocou špeciálnych nástrojov CAD (obr. 3.1-11), ktoré umožňujú podrobne vizualizovať prvky projektu ako súbor čiar (2D) alebo geometrických telies (3D).

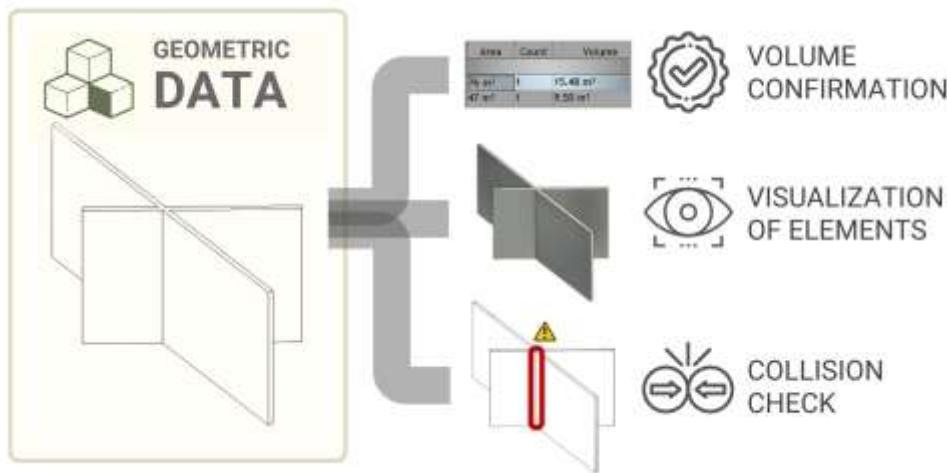


Obr. 3.1-11 Nástroje CAD pomohli preniesť geometrické informácie z fyzických médií do databázovej formy.

Pri práci s geometrickými údajmi v stavebníctve a architektúre možno identifikovať tri hlavné aplikácie geometrických údajov (obrázok 3.1-12):

- **Potvrdenie objemov:** geometrické údaje, vytvorené v rámci programov CAD (BIM) pomocou špeciálnych geometrických jadier, sú potrebné na automatické a presné určenie objemov a rozmerov prvkov projektu. Tieto údaje zahŕňajú automaticky vypočítané plochy, objemy, dĺžky a ďalšie dôležité atribúty potrebné na plánovanie, rozpočtovanie a objednávanie zdrojov a materiálov
- **Vizualizácia projektu:** v prípade akýchkoľvek zmien v projekte umožňuje vizualizácia prvkov automatické generovanie aktualizovaných výkresov v rôznych rovinách. Vizualizácia projektu v počiatocných fázach umožňuje rýchlejšie porozumenie medzi všetkými účastníkmi s cieľom ušetriť čas a zdroje počas procesu výstavby.
- **Kontrola kolízií:** V komplexných stavebných a inžinierskych projektoch, kde je rozhodujúca interakcia viacerých kategórií prvkov (napr. potrubia a stien) bez "geometrických konfliktov", zohráva kontrola kolízií kľúčovú úlohu. Používanie softvéru na zisťovanie kolízií umožňuje aktívne identifikovať potenciálne geometrické konflikty medzi prvkami projektu, čím sa predchádza nákladným chybám počas procesu výstavby.

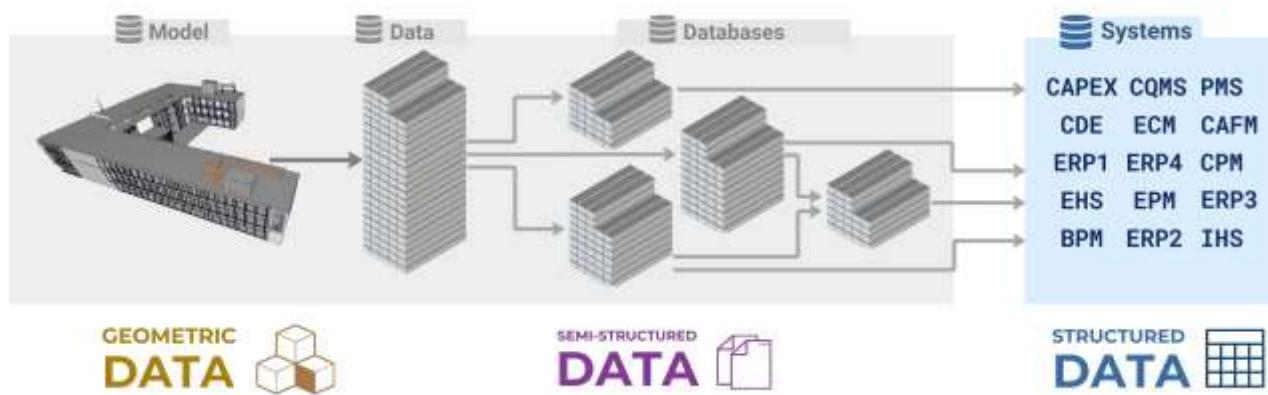
Od samého začiatku existencie inžinierskych projekčných kancelárií, od čias výstavby prvých zložitých stavieb, poskytovali stavební inžinieri geometrické informácie vo forme výkresov, čiar a plošných geometrických prvkov (na papieri, papieri "A0" alebo vo formátoch DWG, PDF, PLT), na základe ktorých majstri a odhadcovia (obr. 3.1-11), posledné tisícročia pomocou pravítok a uhlomerov zhromažďovali atribútové objemy alebo počet prvkov a skupín prvkov. 3.1-11), za posledné tisícročia pomocou pravítok a transportov zbierali atribútové objemy alebo množstvá prvkov a skupín prvkov.



Obr. 3.1-12 Geometria je základom pre získanie objemových parametrov prvkov, ktoré sa potom používajú na výpočet nákladov a časového plánu projektu.

Dnes sa táto manuálna a časovo náročná úloha rieši plnou automatizáciou vďaka nástupu objemového modelovania v moderných nástrojoch CAD (BIM), ktoré umožňuje automaticky, pomocou špeciálneho geometrického jadra, získať objemové atribúty akéhokoľvek prvku bez potreby ručného výpočtu objemových parametrov.

Moderné nástroje CAD umožňujú aj klasifikáciu a kategorizáciu prvkov projektu, aby ste mohli z databázy projektu nahrať tabuľky špecifikácií na použitie v rôznych systémoch, ako je odhadovanie nákladov, plánovanie alebo výpočet CO<sub>2</sub> (obr. 3.1-13). Získavaniu špecifikácií, tabuľiek QTO a množstiev a praktickým príkladom sa venujeme v kapitole "Získavanie množstiev a kvantifikácia".



Obr. 3.1-13 Nástroje CAD (BIM) ukladajú údaje do databáz, ktoré sú určené na integráciu a interakciu s inými systémami.

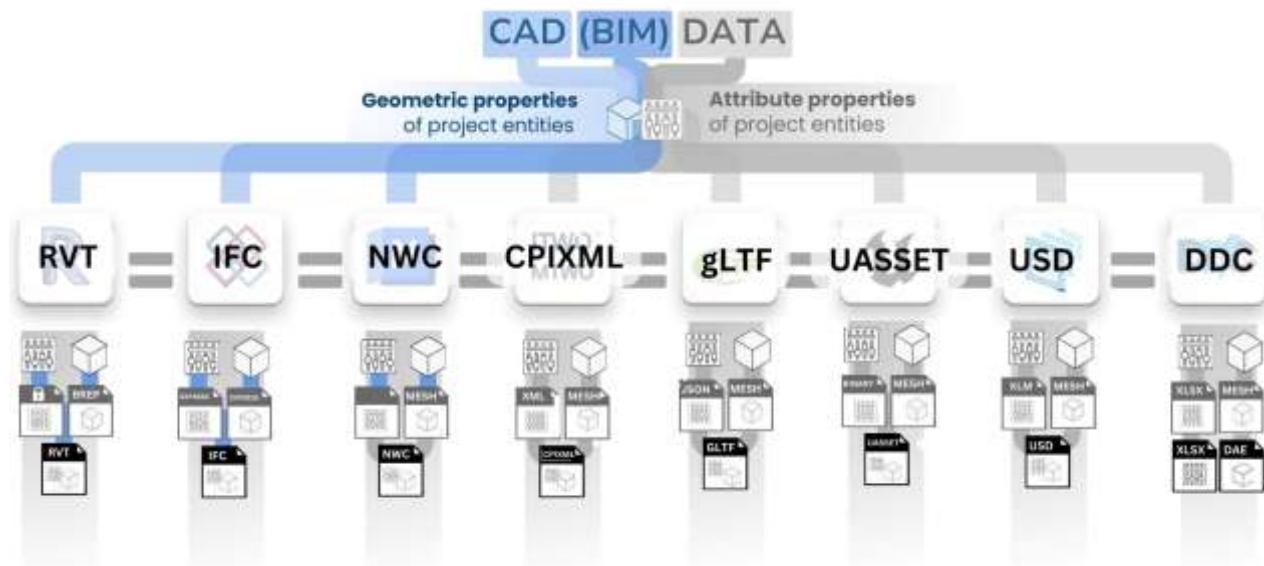
Vzhľadom na uzavretosť databáz a formátov používaných v prostredí CAD sa geometrické údaje vytvorené v riešeniach CAD stali vlastne samostatným typom informácií. Spájajú v sebe geometriu prvkov aj metainformácie (štruktúrované alebo pološtruktúrované), uzavreté v špecializovaných súboroch a formátoch.

## Údaje CAD: od návrhu po ukladanie údajov

Moderné systémy CAD a BIM ukladajú údaje vo vlastných, často proprietárnych formátoch: DWG, DXF, RVT, DGN, PLN a ďalších. Tieto formáty podporujú 2D aj 3D reprezentáciu objektov, pričom zachovávajú nielen geometriu, ale aj atribúty súvisiace s objektmi. Tu sú tie najbežnejšie:

- **DWG** je binárny formát súboru, ktorý sa používa na ukladanie dvojrozmerných (a menej často trojrozmerných) údajov o návrhu a metadát.
- **DXF** je textový formát na výmenu 2D a 3D výkresov medzi CAD systémami. Obsahuje geometriu, vrstvy a atribútové údaje, podporuje ASCII aj binárnu reprezentáciu.
- **RVT** je binárny formát na ukladanie modelov CAD vrátane 3D -geometrie, atribútov prvkov, vzťahov a parametrov návrhu.
- **IFC** je otvorený textový formát na výmenu stavebných údajov medzi systémami CAD (BIM). Obsahuje geometriu, vlastnosti objektov a informácie o ich vzťahoch.

Okrem týchto formátov sa používajú aj ďalšie formáty: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Hoci sa líšia účelom a úrovňou otvorenosti (obrázok 3.1-14), všetky môžu reprezentovať rovnaký informačný model projektu v rôznych formách. V komplexných projektoch sa tieto formáty často používajú súbežne, od výkresu až po koordinačné modely projektu.



Obrázok 3.1-14 Oblúbené formáty ukladania CAD popisujú geometriu prostredníctvom parametrov BREP alebo MESH doplnených atribútmi.

Všetky uvedené formáty umožňujú ukladať údaje o každom prvku stavebného projektu a všetky uvedené formáty obsahujú dva kľúčové typy údajov:

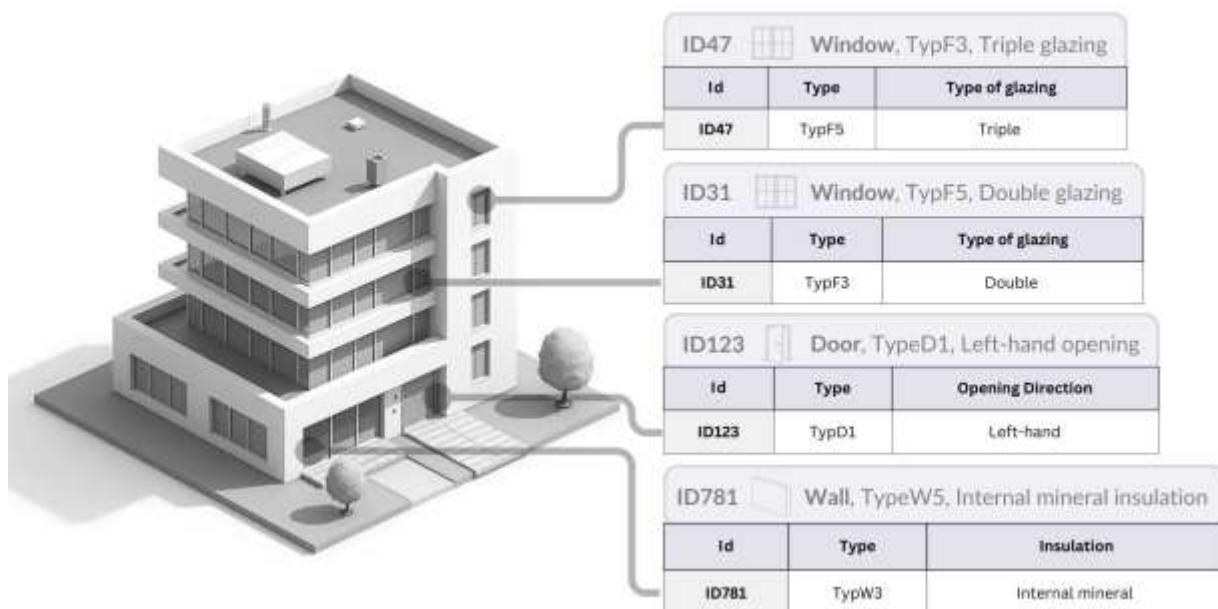
- **Geometrické parametre** - opisujú tvar, polohu a rozmery objektu. Geometriou a jej využitím sa budeme podrobne zaoberať v šiestej časti knihy venovanej riešeniam CAD (BIM);
- **Vlastnosti atribútov** - obsahujú rôzne informácie: materiály, typy prvkov, technické charakteristiky, jedinečné identifikátory a ďalšie vlastnosti, ktoré môžu mať prvky projektu.

Atribútové údaje majú v moderných projektoch mimoriadny význam, pretože definujú prevádzkové charakteristiky objektov, umožňujú inžinierske výpočty a kalkulácie nákladov a zabezpečujú kompletnú interakciu medzi účastníkmi projektovania, výstavby a prevádzky. Napríklad:

- Pre okná a dvere: typ konštrukcie, typ zasklenia, smer otvárania (obr. 3.2-1).
- Pri stenách sa zaznamenávajú informácie o materiáloch, tepelnej izolácii a akustických vlastnostiach.
- V prípade inžinierskych systémov sú uložené parametre potrubí, kanálov, káblových trás a ich prípojok.

Tieto parametre môžu byť uložené v samotných súboroch CAD-(BIM-) a v externých databázach - ako výsledok exportu, konverzie alebo priameho prístupu k interným štruktúram CAD prostredníctvom nástrojov reverzného inžinierstva. Tento prístup uľahčuje integráciu informácií o návrhu s inými podnikovými systémami a platformami

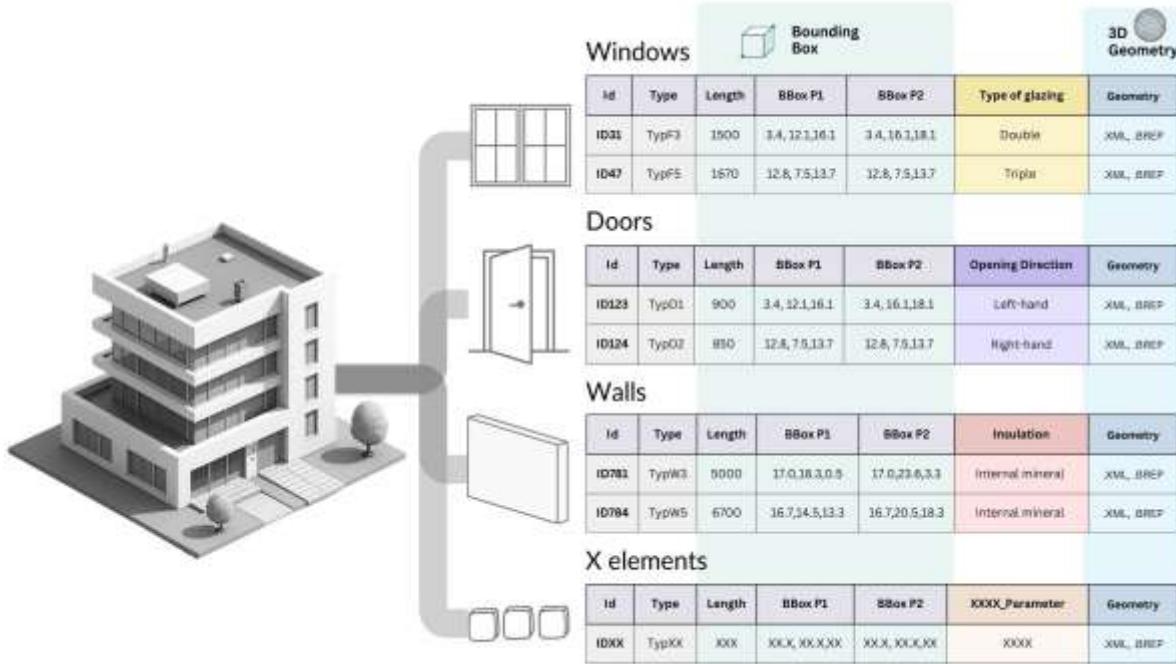
**Reverzné inžinierstvo** v kontexte CAD (BIM) je proces extrakcie a analýzy vnútornej štruktúry digitálneho modelu s cieľom obnoviť jeho logiku, štruktúru údajov a závislosti bez prístupu k pôvodným algoritmom alebo dokumentácii.



Obr. 3.1-15 Projektový prvok okrem popisu parametrickej alebo polygonálnej geometrie obsahuje aj informácie o parametroch a vlastnosťach prvkov.

V dôsledku toho sa okolo každého prvku vytvorí jedinečný súbor parametrov a vlastností, ktorý zahŕňa jedinečné charakteristiky každého objektu (napr. identifikátor a rozmery) a spoločné atribúty pre skupiny prvkov. To umožňuje nielen analyzovať jednotlivé prvky - entity projektu, ale aj spájať ich do logických skupín, ktoré potom môžu využívať iní špecialisti na svoje úlohy a výpočty v systémoch a databázach.

Entita je konkrétny alebo abstraktný objekt reálneho sveta, ktorý možno jednoznačne identifikovať, opísť a reprezentovať vo forme údajov.

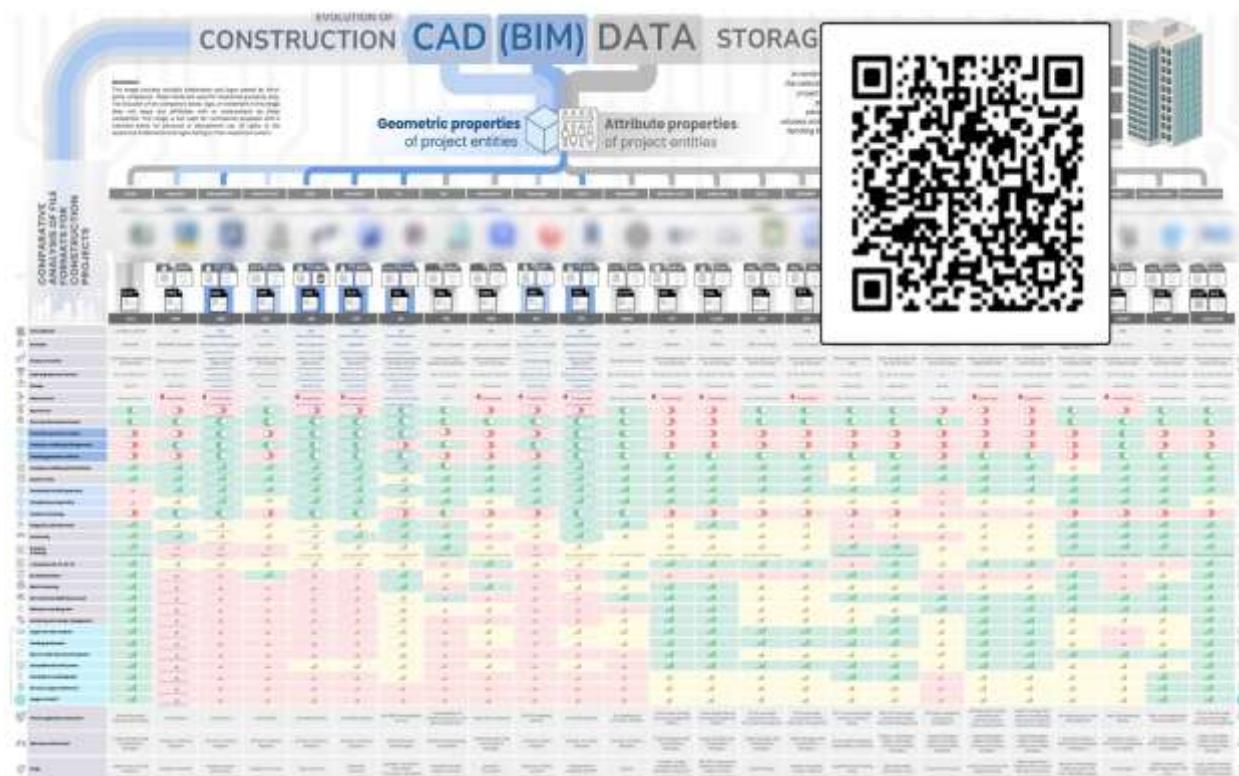


Obr. 3.1-16 Každý prvok projektu obsahuje atribúty, ktoré bud' zadáva projektant, alebo sú vypočítané v programe CAD.

V posledných desaťročiach sa v stavebníctve vyvinulo mnoho nových formátov CAD (BIM), ktoré zjednodušujú tvorbu, ukladanie a prenos údajov. Tieto formáty môžu byť uzavreté alebo otvorené, tabuľkové, parametrické alebo grafické. Ich rôznorodosť a roztrieštenosť však výrazne komplikuje správu údajov vo všetkých fázach životného cyklu projektu. Porovnávacia tabuľka hlavných formátov používaných na výmenu informácií v stavebníctve je uvedená na obrázku 3.1-17 (plná verzia je k dispozícii prostredníctvom QR kódu).

Na riešenie problémov interoperability a prístupu k údajom CAD sa podieľajú manažéri BIM (BIM) a koordinátori, ktorých úlohou je kontrolovať export, kontrolovať kvalitu údajov a integrovať časti údajov CAD (BIM) do iných systémov.

Vzhľadom na uzavretosť a zložitosť formátov je však ťažké tento proces automatizovať, čo nútí odborníkov vykonávať mnohé operácie ručne bez možnosti vytvoriť plnohodnotné procesy spracovania údajov v línii (pipeline).



The screenshot shows a complex CAD (BIM) software interface. At the top, there's a navigation bar with tabs like 'CONSTRUCTION', 'CAD (BIM)', 'DATA', and 'STORAGE'. Below the navigation bar is a large table with two main sections: 'Geometric properties of project entities' and 'Attribute properties of project entities'. The table is filled with numerous rows and columns of data, each representing a different entity or component within a project. On the far left of the table, there's a vertical column of icons and labels related to 'COMPARATIVE ANALYSIS OF PROJECTS'. To the right of the table, there's a small 3D model of a building and a QR code.

Obr. 3.1-17 Tabuľka porovnávajúca hlavné formáty údajov, v ktorých sú uložené informácie o prvkoch projektu [53].

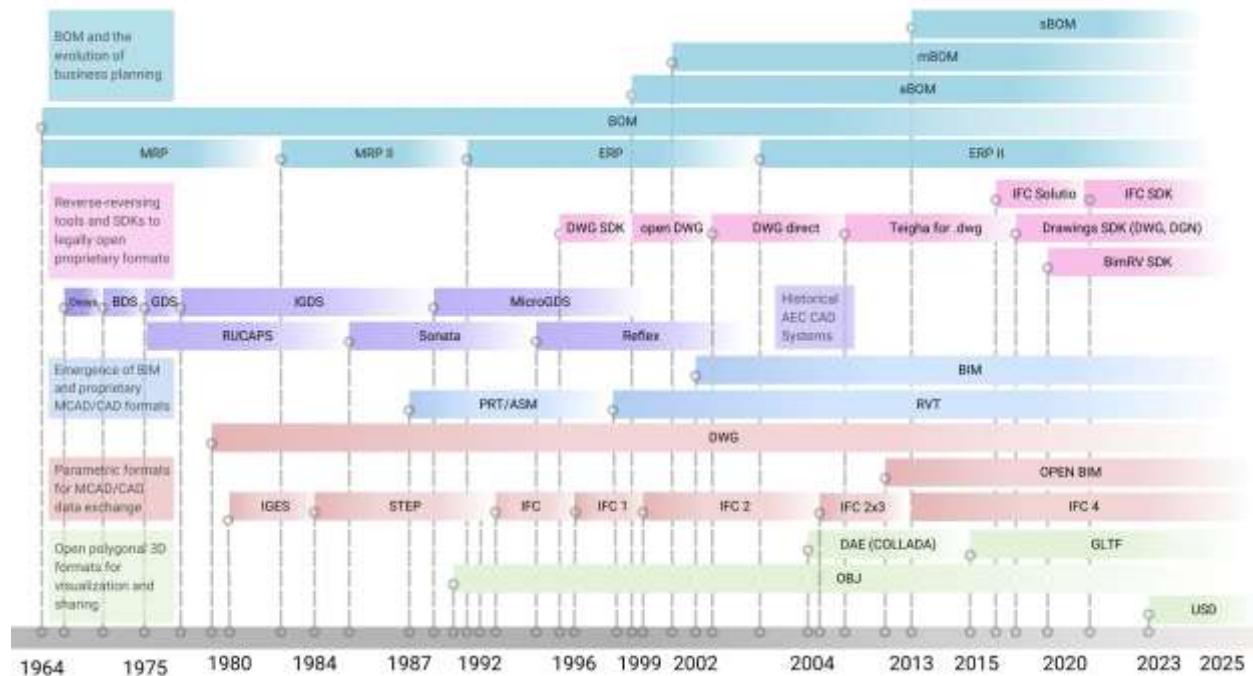
Aby sme pochopili, prečo existuje toľko rôznych dátových formátov a prečo je väčšina z nich uzavretá, je dôležité preniknúť do procesov, ktoré prebiehajú vo vnútri programov CAD (BIM, ktoré budú podrobne preskúmané v šiestej časti knihy).

Ďalšiu informačnú vrstvu pridanú ku geometrii zaviedli vývojári CAD systémov v podobe konceptu BIM (Building Information Modeling), marketingového pojmu, ktorý sa v stavebníctve aktívne presadzuje od roku 2002 [54].

## Vznik koncepcie BIM (BOM) a využívanie CAD v procesoch

Koncepcia informačného modelovania budov (BIM), prvýkrát načrtnutá v dokumente BIM Whitepaper z roku 2002 [54], vznikla na základe marketingových iniciatív výrobcov softvéru CAD. Vznikla z marketingových iniciatív výrobcov softvéru CAD a bola pokusom prispôsobiť princípy, ktoré už boli dobre zavedené v strojárstve, potrebám stavebného priemyslu.

Inšpiráciou pre BIM bol koncept BOM (Bill of Materials) - kusovník, ktorý sa v priemysle hojne využíva od konca 80. rokov minulého storočia. V strojárstve BOM umožnil prepojenie údajov zo systémov CAD so systémami PDM (Product Data Management), PLM (Product Lifecycle Management) a ERP, čím sa zabezpečila komplexná správa technických informácií počas celého životného cyklu výrobku (obr. 3.1-8).



Obr. 3.1-18 Vývoj špecifikácií (BOM), informačného modelovania (BIM) a digitálnych formátov v inžinierskom stavebníctve.

Moderný vývoj koncepcie BOM viedol k vzniku rozšíreného rámca - XBOM (Extended BOM), ktorý zahŕňa nielen zloženie výrobku, ale aj scenáre správania, prevádzkové požiadavky, parametre udržateľnosti a údaje pre prediktívnu analýzu. XBOM v podstate plní rovnakú úlohu ako BIM v stavebníctve: oba prístupy sa snažia premeniť digitálny model na jednotný zdroj pravdy pre všetkých účastníkov projektu počas celého životného cyklu projektu.

Kľúčovým miľníkom pri vzniku BOM v stavebníctve bolo zavedenie prvého parametrického CAD (MCAD) špeciálne prispôsobeného pre stavebníctvo v roku 2002. Vyvinul ho tím, ktorý predtým vytvoril Pro-E®, revolučný systém MCAD pre strojárstvo, ktorý vznikol koncom 80. rokov a stal sa priemyselným štandardom [55].

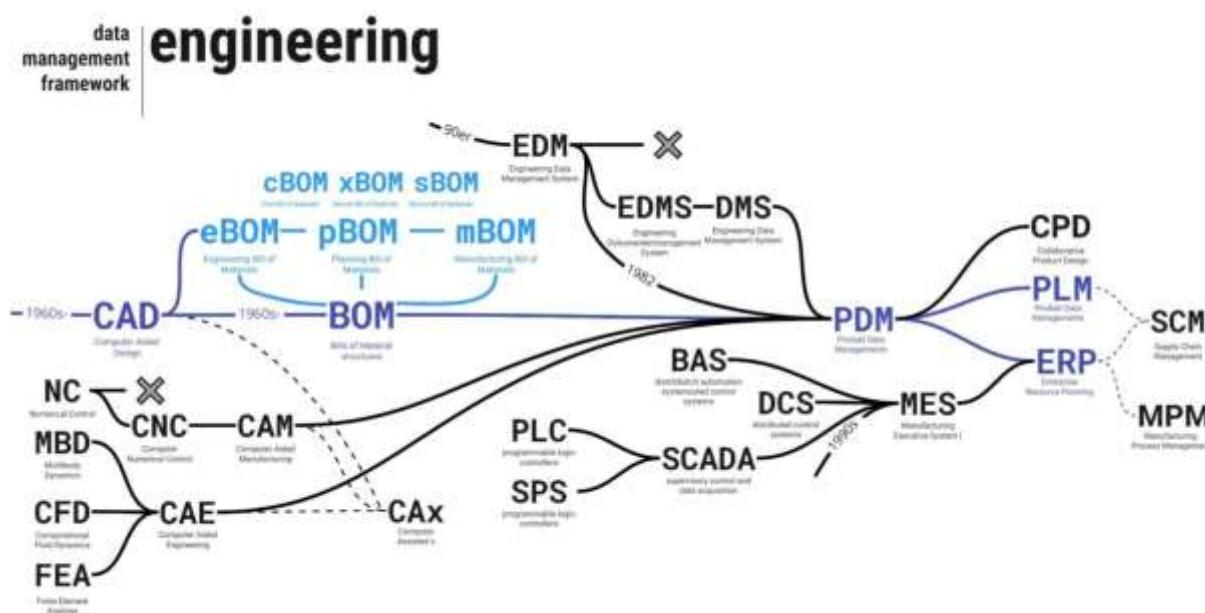
Už koncom 80. rokov bolo cieľom odstrániť obmedzenia [56] vtedy existujúcich CAD -programov. Hlavným cieľom bolo znížiť prácnosť pri vykonávaní zmien parametrov konštrukčných prvkov a umožniť aktualizáciu modelu na základe údajov mimo CAD programov prostredníctvom databázy [57]. Najdôležitejšiu úlohu pritom mala zohrávať parametrizácia: automatické získavanie charakteristík z databázy a ich využívanie na aktualizáciu modelu v rámci systémov CAD.

Pro-E a koncepcia elementárneho parametrického modelovania c BOM, ktorá je jeho základom, mali významný vplyv na rozvoj trhu CAD - a MCAD [58]. Tento model je v priemysle už 25 rokov a mnogé moderné systémy sa stali jeho koncepcnými nasledovníkmi.

Cieľom je vytvoriť systém, ktorý je dostatočne flexibilný na to, aby mohol inžinier ľahko zvážiť rôzne návrhy. A náklady na vykonanie zmien v návrhu by mali byť čo najbližšie k nule. Tradičný softvér CAD / CAM nerealisticky obmedzuje vykonávanie lacných zmien len na samom začiatku procesu návrhu [59].

- Samuel Heisenberg, zakladateľ spoločnosti Parametric Technology Corporation®, vývojár MCAD -produkту Pro-E a učiteľ tvorca CAD produktu využívajúceho formát RVT

V strojárstve sa klúčovými platformami stali systémy PDM, PLM, MRP a ERP. Zohrávajú ústrednú úlohu pri správe údajov a procesov, zhromažďujú informácie z CAx systémov (CAD, CAM, CAE) a organizujú konštrukčné činnosti na základe štruktúry výrobku (BOM: eBOM, pBOM, mBOM) (obr. 3.1-18). Táto integrácia znižuje počet chýb, zabraňuje duplike údajov a zabezpečuje komplexnú sledovateľnosť od návrhu až po výrobu.



Obrázok 3.1-19 Historicky sa BOM objavil v 60. rokoch 20. storočia ako spôsob štruktúrovania údajov zo systémov CAx a ich odovzdávania riadiacim systémom.

Zakúpenie riešenia CAD vyvinutého bývalým tímom Pro-E a založeného na prístupe BOM jedným z popredných dodávateľov bolo takmer okamžite poznačené uverejnením série bielych kníh BIM (2002-2003)[60][61]. Už od polovice roku 2000 sa koncepcia BIM začala aktívne presadzovať v stavebníctve, čo výrazne zvýšilo záujem o parametrický softvér. Popularita rástla tak rýchlo, že stavebná vidlica strojárskeho softvéru Pro-E - parametrický CAD propagovaný týmto výrobcom - skutočne vytlačila konkurenciu v segmente architektonického a konštrukčného navrhovania (obr. 3.1-20). Začiatkom roka 2020 si de facto upevnil globálnu dominanciu na trhu BIM (CAD) [62].



Obr. 3.1-20 Popularita vyhľadávacích dotazov Google (RVT verus IFC): parametrický CAD vytvorený bývalým tímom Pro-E s podporou BOM -BIM získal popularitu takmer vo väčšine krajín sveta.

Za posledných 20 rokov získala skratka BIM množstvo výkladov, ktorých mnohomôžnosť má korene v pôvodných marketingových koncepciach, ktoré vznikli začiatkom roku 2000. Norma ISO 19650, ktorá zohrala dôležitú úlohu pri popularizácii tohto pojmu, vlastne zabezpečila status BIM ako "vedecky podloženého" prístupu k riadeniu informácií. V samotnom teste normy, ktorá sa venuje správe údajov počas celého životného cyklu objektov pomocou BIM, sa však skratka BIM spomína, ale nikdy nie je jasne definovaná

Pôvodná webová stránka dodávateľa, ktorá v rokoch 2002 [60] a 2003 [61] uverejnila sériu bielych kníh o BIM, v skutočnosti reprodukovala marketingové materiály o koncepciach BOM (Bills of Materials) a PLM (Product Lifecycle Management), ktoré sa predtým používali v strojárskom softvéri Pro-E v 90. rokoch [63].

*Informačné modelovanie budov, inovatívny nový prístup k navrhovaniu, výstavbe a riadeniu budov, ktorý v roku 2002 predstavila spoločnosť..... [názov dodávateľa CAD], zmenil spôsob, akým odborníci na celom svete uvažujú o tom, ako možno technológiu použiť pri navrhovaní, výstavbe a riadení budov.*

- Biela kniha BIM, 2003 [61].

Tieto prvé publikácie priamo spájali BIM s koncepciou centralizovanej integrovanej databázy. Ako sa uvádzá v Bielej knihe z roku 2003, BIM je správa informácií o budove, pri ktorej sa všetky aktualizácie uskutočňujú v jedinom úložisku, čím sa zabezpečí synchronizácia všetkých výkresov, rezov a špecifikácií (BOM - Bills of Materials).

*BIM sa opisuje ako správa informácií o budove, kde sa všetky aktualizácie a zmeny uskutočňujú v databáze. Či už teda pracujete so schémami, rezmi alebo výkresmi listov, všetko je vždy koordinované, konzistentné a aktuálne.*

- Webová stránka spoločnosti CAD vendor with BIM Whitepaper, 2003 [54].

Myšlienka riadenia dizajnu prostredníctvom jednej integrovanej databázy bola široko diskutovaná už vo výskume v 80. rokoch 20. storočia. Napríklad koncepcia BDS Charlesa Eastmana [57] obsahovala 43 odkazov na pojem "databáza" (obrázok 6.1-2). Do roku 2004 sa tento počet znížil takmer o polovicu na 23 v Bielej knihe o BIM z roku 2002 [64]. A do polovice roku 2000 sa téma databáz prakticky vytratila z marketingových materiálov dodávateľov a z agendy digitalizácie vo všeobecnosti.

Hoci pôvodne bola jadrom systému BIM práve databáza a prístup k nej, časom sa dôraz presunul na geometriu, vizualizáciu a 3D. Samotný registrátor štandardu IFC v roku 1994, ktorý v roku 2002 vydal Bielu knihu BIM - ten istý dodávateľ - v Bielej knihe zo začiatku roka 2000 výslovne poukázal na obmedzenia neutrálnych formátov, ako sú IGES, STEP a IFC, a na potrebu priameho prístupu k databázam CAD:

*Rôzne aplikácie môžu byť nekompatibilné a opäťovne zadané údaje môžu byť nepresné [...]. Výsledok tradičného počítačom podporovaného navrhovania [CAD]: vyššie náklady, dlhší čas na uvedenie na trh a nižšia kvalita výrobku. V súčasnosti všetky hlavné aplikácie používajú štandardné priemyselné rozhrania na výmenu údajov na nízkej úrovni. Používaním starých štandardov IGES alebo nových štandardov STEP [IFC je de facto a de jure kópiou formátu STEP/IGES] na výmenu údajov medzi aplikáciami od rôznych výrobcov môžu používateľia dosiahnuť určitú kompatibilitu údajov medzi najlepšími výrobkami. IGES a STEP však fungujú len na nízkych úrovniach a nedokážu si vymieňať tak bohaté údaje, ako sú informácie generované dnešnými špičkovými aplikáciami [...]. A hoci sa tieto a ďalšie normy takmer denne zlepšujú, z hľadiska bohatosti údajov budú vždy zaostávať za produktmi dnešných výrobcov. [...] Programy v rámci aplikácie musia byť schopné vymieňať a zachovať bohatosť údajov bez toho, aby sa uchýlili k neutrálnym prekladačom, ako sú IGES, STEP [IFC] alebo PATRAN. Namiesto toho by rámcové aplikácie mali mať možnosť priameho prístupu k základnej databáze CAD, aby sa nestratila podrobnosť a presnosť informácií.*

- Biela kniha dodávateľa CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale" (Integrovaný návrh a výroba: výhody a odôvodnenie), 2000 [65].

Preto sa už v 80. rokoch 20. storočia a začiatkom 21. storočia za kľúčový prvok digitálneho navrhovania v prostredí CAD považovala skôr databáza než formátovací súbor alebo neutrálny formát IFC. Navrhovalo sa, aby sa upustilo od prekladačov a aby aplikácie mali priamy prístup k údajom. V skutočnosti však v polovici 20. rokov 20. storočia koncepcia BIM začala pripomínať strategiu "rozdeľuj

a panuj", v ktorej sa uprednostňujú záujmy dodávateľov softvéru využívajúcich uzavreté geometrické jadrá pred rozvojom otvorenej výmeny informácií.

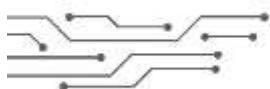
BIM sa dnes vníma ako neoddeliteľná súčasť stavebníctva. Za posledné dve desaťročia sa však prísluby zjednodušenej spolupráce a integrácie údajov zväčša nenaplnili. Väčšina riešení je stále viazaná na uzavreté alebo neutrálne formáty a špecializované nástroje. Histórii BIM, otvorenému BIM a IFC, ako aj otázkam interoperability a geometrických jadier sa podrobne venujeme v 6. časti knihy "CAD a BIM: Marketing, realita a budúcnosť projektových údajov v stavebníctve".

V súčasnosti stojí priemysel pred kľúčovou výzvou prejsť od tradičného chápania CAD (BIM) ako nástroja na modelovanie k jeho využívaniu ako plnohodnotnej databázy. To si vyžaduje nové prístupy k práci s informáciami, opustenie závislosti od uzavretých ekosystémov a zavedenie otvorených riešení.

S rozvojom nástrojov reverzného inžinierstva, ktoré umožňujú prístup k databázam CAD, ako aj s rozšírením technológií Open Source a LLM sa používatelia a vývojári v stavebníctve čoraz viac odkláňajú od nejasných podmienok dodávateľov softvéru. Namiesto toho sa pozornosť presúva na to, čo je skutočne dôležité: údaje (databázy) a procesy.

Za módnymi skratkami a vizualizáciami sa skrývajú štandardné postupy správy údajov: ukladanie, prenos a transformácia - t. j. klasický proces ETL (Extract, Transform, Load). Podobne ako v iných odvetviach si digitalizácia stavebníctva vyžaduje nielen výmenné štandardy, ale aj jasne štruktúrované spracovanie heterogénnych informácií.

Aby sa naplno využil potenciál údajov CAD (BIM), musia spoločnosti prehodnotiť svoj prístup k správe informácií. To nevyhnutne povedie ku kľúčovému prvku digitálnej transformácie - zjednoteniu, štandardizácií a zmysluplnému štruktúrovaniu údajov, s ktorými stavební odborníci denne pracujú.

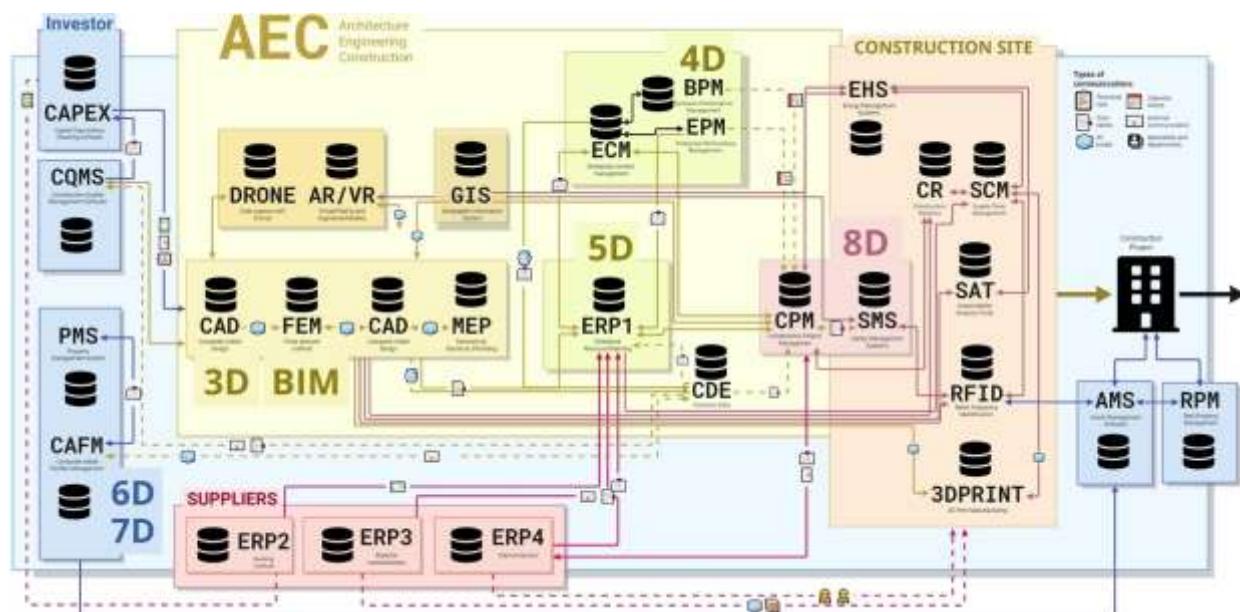


## KAPITOLA 3.2.

# HARMONIZÁCIA A ŠTRUKTÚROVANIE ÚDAJOV

### Napĺňanie systémov údajmi v stavebníctve

Či už ide o veľké korporácie alebo stredne veľké podniky, špecialisti sa denne zaoberajú napĺňaním softvérových systémov a databáz s rôznymi rozhraniami multiformátovými informáciami (obr. 3.2-1), ktoré musia s pomocou manažérov navzájom spolupracovať. Práve tento komplex vzájomne sa ovplyvňujúcich systémov a procesov v konečnom dôsledku vytvára príjmy a zisk spoločnosti.



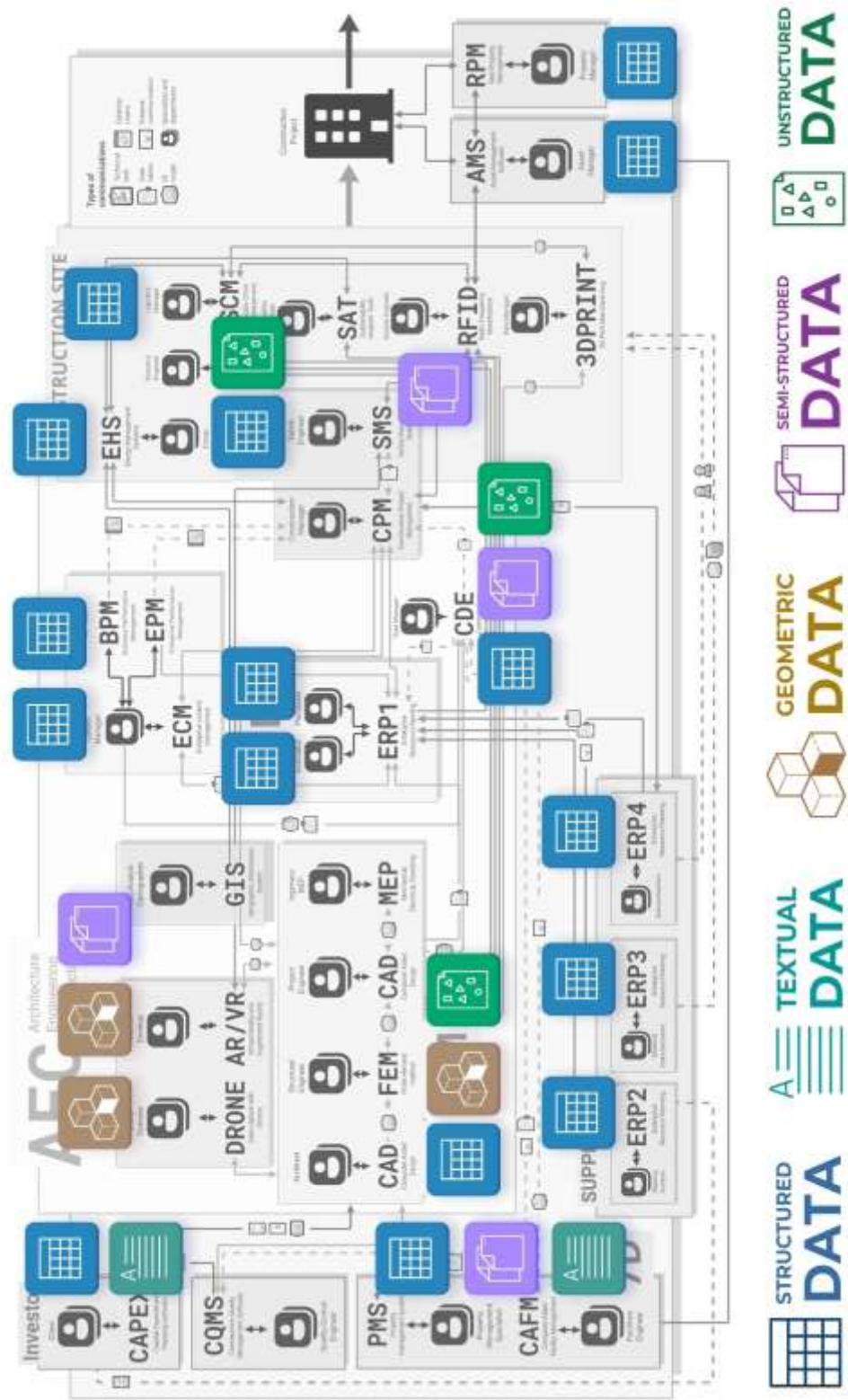
Obrázok 3.2-1 Prakticky každý systém alebo aplikácia používaná v stavebníctve má vo svojom jadre niektorú z populárnych databáz RDBMS.

Každá z uvedených kategórií systémov, ktoré sa uplatňujú v stavebníctve, pracuje s vlastnými dátovými typmi zodpovedajúcimi funkčnej úlohe týchto systémov. Aby sme prešli z abstraktnej úrovne na konkrétnu, prejdeme od dátových typov k ich reprezentácii vo forme formátov a dokumentov.

K predtým uvedenému zoznamu systémov (obrázok 1.2-4) teraz pridávame konkrétnie typy formátov a dokumentov, s ktorými často pracujú:

#### Investor (CAPEX)

- Finančné údaje: rozpočty, prognózy výdavkov (štruktúrované údaje).
- Údaje o trendoch na trhu: analýzy trhu (štruktúrované a neštruktúrované údaje).
- Právne a zmluvné údaje: zmluvy (textové údaje).



Obrázok 3.2-2 V stavebníctve sa používa mnoho systémov s rôznymi rozhraniami, ktoré pracujú s rôznymi typmi údajov.

### **Systémy riadenia (PMS, CAFM, CQMS)**

- Údaje o projekte: grafy, úlohy (štruktúrované údaje).
- Údaje o údržbe zariadení: plány údržby (textové a pološtruktúrované údaje).
- Údaje o kontrole kvality: normy, kontrolné správy (textové a neštruktúrované údaje).

### **CAD, FEM a BIM**

- Technické výkresy: architektonické, konštrukčné plány (geometrické údaje, neštruktúrované údaje).
- Stavebné modely: 3D -modely, materiálové údaje (geometrické a pološtruktúrované údaje).
- Technické výpočty: analýza zaťaženia (štruktúrované údaje).

### **Systémy riadenia staveniska (EHS, SCM)**

- Údaje o bezpečnosti a zdraví: bezpečnostné protokoly (textové a štruktúrované údaje).
- Údaje o dodávateľskom reťazci: zásoby, objednávky (štruktúrované údaje).
- Denné správy: pracovný čas, produktivita (štruktúrované údaje).

### **Drony, AR/VR, GIS, 3D - tlač**

- Geodáta: topografické mapy (geometrické a štruktúrované údaje).
- Údaje v reálnom čase: video a fotografie (neštruktúrované údaje).
- Modely pre 3D -tlač: digitálne výkresy (geometrické údaje).

### **Ďalšie systémy riadenia (4D BPM, 5D ERP1)**

- Údaje o čase a nákladoch: harmonogramy, odhady (štruktúrované údaje).
- Riadenie zmien: záznamy o zmenách projektu (textové a štruktúrované údaje).
- Vykazovanie výkonnosti: ukazovatele úspešnosti (štruktúrované údaje).

### **Integrácia údajov a komunikácia (CDE, RFID, AMS, RPM)**

- Výmena údajov: výmena dokumentov, dátové modely (štruktúrované a textové údaje).
- RFID a sledovanie údajov: logistika, správa majetku (štruktúrované údaje).
- Monitorovanie a kontrola: senzory na miestach (štruktúrované a neštruktúrované údaje).

Každý systém v stavebnictve - od systémov riadenia staveniska až po prevádzkové databázy - pracuje s vlastným typom informácií: štruktúrovanými, textovými, geometrickými a inými. "Dátové prostredie", s ktorým musia odborníci denne pracovať, je veľmi rôznorodé. Jednoduchý výpočet formátov však neodhaluje zložitosť skutočnej práce s informáciami.

V praxi sa spoločnosti stretávajú s tým, že údaje, aj keď sú získané zo systémov, nie sú pripravené na použitie "tak, ako sú". Týka sa to najmä textov, obrázkov, súborov PDF, súborov CAD a iných formátov, ktoré je ľahké analyzovať pomocou štandardných nástrojov. Preto je ďalším kľúčovým krokom transformácia údajov - proces, bez ktorého nie je možné efektívne automatizovať spracovanie, analýzu, vizualizáciu a rozhodovanie.

## **Transformácia údajov: rozhodujúci základ modernej podnikovej analýzy**

Väčšina spoločností dnes čeli paradoxu: približne 80% ich každodenných procesov sa stále spolieha na klasické štruktúrované údaje - známe tabuľky Excelu a relačné databázy (RDBMS) [66]. Zároveň

však 80% nových informácií, ktoré vstupujú do digitálneho ekosystému spoločnosti, je neštruktúrovaných alebo voľne štruktúrovaných (obr. 3.2-3) [52]. Patrí sem text, grafika, geometria, obrázky, CAD -modely, dokumentácia vo formáte PDF, zvukové a obrazové záznamy, elektronická korešpondencia a mnohé ďalšie.

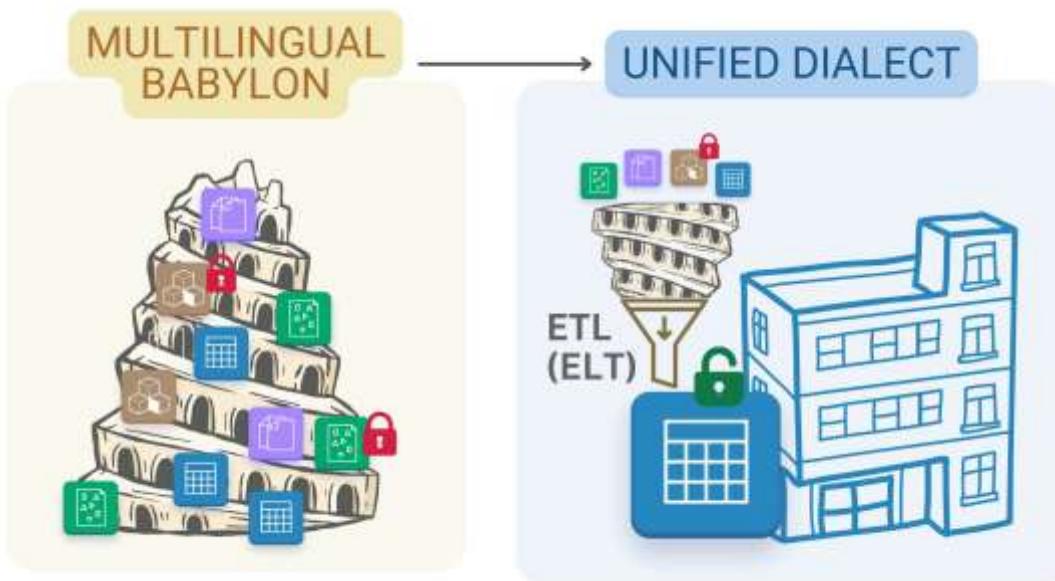
Okrem toho objem neštruktúrovaných údajov naďalej rýchlo rastie - ročné tempo rastu sa odhaduje na 55-65% [67]. Takáto dynamika spôsobuje vážne problémy pri integrácii nových informácií do existujúcich podnikových procesov. Ignorovanie tohto toku viacformátových údajov vedie k vzniku informačných medzier a znížuje spravovateľnosť celého digitálneho prostredia podniku.



Obr. 3.2-3 Každoročný nárast neštruktúrovaných údajov vytvára výzvy pri integrácii prúdových informácií do podnikových procesov.

Ignorovanie zložitých neštruktúrovaných a mätúcich voľne štruktúrovaných údajov v procesoch automatizácie môže viesť k výrazným medzerám v informačnom prostredí spoločnosti. V dnešnom svete nekontrolovateľného a lavínovitého pohybu informácií si spoločnosti musia osvojiť hybridný prístup k správe údajov, ktorý zahŕňa účinné metódy na prácu so všetkými typmi údajov.

Klúčom k efektívnej správe údajov je organizácia, štruktúrovanie a klasifikácia rôznych typov údajov "Babel" (vrátane neštruktúrovaných, textových a geometrických formátov, do štruktúrovaných alebo voľne štruktúrovaných údajov). Tento proces transformuje chaotické súbory údajov do organizovaných štruktúr na integráciu do systémov, čím umožňuje rozhodovanie na ich základe (obrázok 3.2-4).



Obr. 3.2-4 Hlavnou úlohou oddelení správy údajov je previesť "Babylon" rôznorodých a mnohoformátových údajov do štruktúrovaného a kategorizovaného systému.

Jednou z hlavných prekážok takejto harmonizácie zostáva nízka úroveň interoperability medzi rôznymi digitálnymi platformami - "silá", o ktorých sme hovorili v predchádzajúcich kapitolách.

Národný inštitút pre štandardy a technológie (NIST, USA) podľa správy zdôrazňuje [68], že nedostatočná kompatibilita údajov medzi rôznymi platformami budov vedie k strate informácií a značným dodatočným nákladom. Len v roku 2002 spôsobili problémy s interoperabilitou softvéru straty v investičnej výstavbe v USA v celkovej výške 15,8 miliardy USD ročne, pričom dve tretiny týchto strát znásajú vlastníci a prevádzkovatelia budov, najmä počas prevádzky a údržby [68]. V štúdiu sa tiež uvádzá, že štandardizácia dátových formátov môže znížiť tieto straty a zvýšiť efektívnosť počas celého životného cyklu objektu.

Podľa štúdie CrowdFlower z roku 2016 [69], ktorá sa týkala 16 000 dátových vedcov z celého sveta, hlavným problémom zostávajú "špinavé" a multiformátové údaje. Podľa tejto štúdie najcennejším zdrojom nie sú konečné databázy alebo modely strojového učenia, ale čas strávený prípravou informácií.

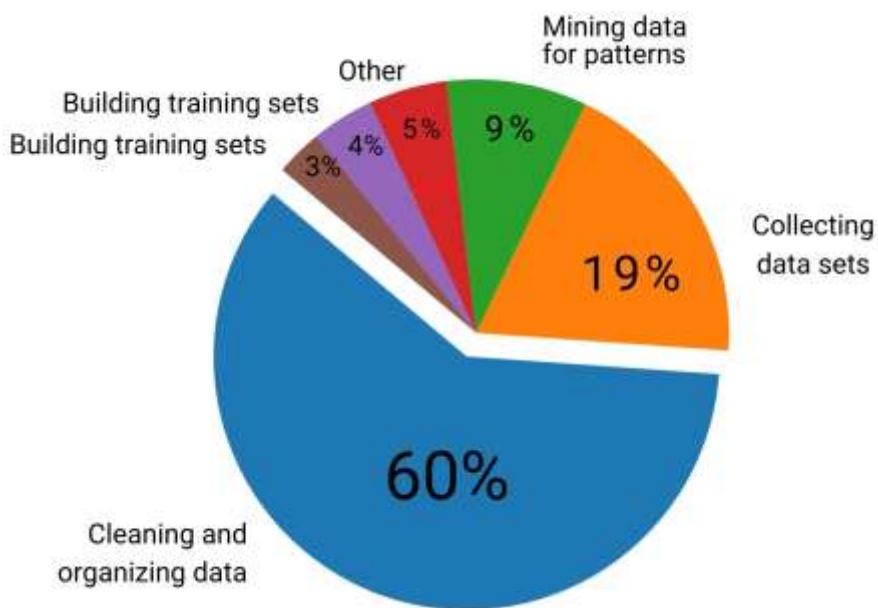
Čistenie, formátovanie a organizovanie zaberá až 60 percent času analytika a správcu údajov. Takmer pätnať strávi hľadaním a zhromažďovaním správnych súborov údajov, ktoré sú často ukryté v silách a nedostupné pre analýzu. A len približne 9 percent času sa venuje priamo modelovaniu, analýze, vytváraniu predpovedí a testovaniu hypotéz. Zvyšok sa venuje komunikácii, vizualizácii, reportovaniu a vyhľadávaniu podporných informačných zdrojov.

V priemere je práca manažéra s údajmi rozdelená takto (obrázok 3.2-5):

- **Čistenie a organizácia údajov (60%):** Čisté a štruktúrované údaje môžu výrazne skrátiť pracovný čas analytika a urýchliť proces plnenia úloh.
- **Zhromažďovanie údajov (19%):** Hlavnou výzvou pre odborníkov na vedu o údajoch je nájsť

relevantné súbory údajov. Firemné údaje sú často uložené v chaoticky usporiadaných "silách", čo stáže prístup k potrebným informáciám.

- **Modelovanie/strojové učenie (9%):** Často je problémom nedostatočná zrozumiteľnosť obchodných cieľov zo strany zákazníkov. Nedostatok jasného poslania môže negovať potenciál aj toho najlepšieho modelu.
- **Ostatné úlohy (5%):** okrem spracovania údajov sa analytici musia zaoberať výskumom, skúmaním údajov z rôznych hľadísk, oznamovaním výsledkov prostredníctvom vizualizácií a správ a odporúčaním optimalizácie procesov a stratégií.



Obr. 3.2-5 Čomu venujú manažéri pracujúci s údajmi najviac času (podľa [70]).

Tieto odhady potvrdzujú aj ďalšie štúdie. Podľa štúdie Xplenty uverejnenej v časopise BizReport v roku 2015 [71] sa 50% až 90% času profesionálov v oblasti business intelligence (BI) (BI) venuje príprave údajov na analýzu.

Čistenie, validácia a organizácia údajov predstavuje kritický základ pre všetky nadväzujúce dátové a analytické procesy, pričom zaberá až 90% času dátových vedcov.

Táto náročná práca, ktorá je pre koncového používateľa neviditeľná, je veľmi dôležitá. Chyby v prvotných údajoch nevyhnutne skresľujú analýzy, sú zavádzajúce a môžu viesť k nákladným chybám v riadení. Preto sa procesy čistenia a štandardizácie údajov - od odstránenia duplicit a doplnenia chýbajúcich údajov až po harmonizáciu merných jednotiek a zosúladenie so spoločným modelom - stávajú základným kameňom modernej digitálnej stratégie.

Dôkladná transformácia, čistenie a štandardizácia údajov tak nielenže zaberá väčšinu času odborníkov (až 80% práce s údajmi), ale určuje aj možnosť ich efektívneho využitia v rámci moderných podnikových procesov. Samotnou organizáciou a čistením údajov sa však úloha optimálneho riadenia informačných tokov v podniku nevyčerpáva. Vo fáze organizácie a štruktúrovania sa stáva dôležitým

výber vhodného dátového modelu, ktorý priamo ovplyvňuje pohodlie a efektívnosť práce s informáciami v ďalších fázach spracovania.

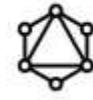
Kedže údaje a obchodné ciele sa líšia, je dôležité pochopíť vlastnosti dátových modelov a vedieť vybrať alebo vytvoriť správnu štruktúru. V závislosti od stupňa štruktúrovania a spôsobu opisu vzťahov medzi prvkami existujú tri hlavné modely: štruktúrovaný, voľne štruktúrovaný a grafický. Každý z nich je vhodný na rôzne úlohy a má svoje silné a slabé stránky.

## Dátové modely: vzťahy v údajoch a vzťahy medzi prvkami

Údaje v informačných systémoch sú organizované rôznymi spôsobmi - v závislosti od úloh a požiadaviek na ukladanie, spracovanie a prenos informácií. Kľúčovým rozdielom medzi typmi dátových modelov, teda formou, v ktorej sú informácie uložené, je stupeň štruktúrovania a spôsob, akým sú opísané vzťahy medzi prvkami.

Štruktúrované údaje majú jasné a opakovateľné schému: sú usporiadane ako tabuľky s pevne stanovenými stĺpcami. Tento formát poskytuje predvídateľnosť, jednoduchosť spracovania a efektívnosť pri vykonávaní dotazov SQL, filtrovaní a agregácií. Príklady - databázy (RDBMS), Excel, CSV.

Voľne štruktúrované údaje umožňujú flexibilnú štruktúru: rôzne prvky môžu obsahovať rôzne sady atribútov a môžu byť uložené ako hierarchie. Príkladom sú formáty JSON, XML alebo iné formáty dokumentov. Takéto údaje sú výhodné, keď je potrebné modelovať vnorené objekty a vzťahy medzi nimi, ale na druhej strane komplikujú analýzu a štandardizáciu údajov (obr. 3.2-6).

	Data Model	Storage Format	Example
	<b>Relational</b>	CSV, SQL	A table of doors in Excel
	<b>Hierarchical</b>	JSON, XML	Nested door objects inside a room
	<b>Graph-based</b>	RDF, GraphDB	Relationships between building elements

Obrázok 3.2-6 Dátový model je logická štruktúra, ktorá opisuje, ako sú údaje v systéme organizované, uložené a spracované.

Výber vhodného formátu závisí od cieľov:

- Ak je dôležitá rýchlosť filtrovania a analýzy - postačia relačné tabuľky (SQL, CSV, RDBMS, stĺpcové databázy).
- Ak sa vyžaduje flexibilita štruktúry, je lepšie použiť JSON alebo XML.

- Ak majú údaje zložité vzťahy - grafové databázy poskytujú prehľadnosť a škálovateľnosť.

V klasických relačných databázach (RDBMS) je každá entita (napr. dvere) reprezentovaná riadkom a jej vlastnosti stĺpcami tabuľky. Napríklad tabuľka položiek z kategórie "Dvere" môže obsahovať polia ID, Výška, Šírka, Požiarna odolnosť a ID miestnosti označujúce miestnosť (obrázok 3.2-7).

V klasických relačných databázach (RDBMS) sa vzťahy vytvárajú vo forme tabuľiek, kde každý záznam predstavuje objekt a stĺpce predstavujú jeho parametre. V tabuľkovom formáte vyzerajú údaje o dverách v projekte takto, pričom každý riadok predstavuje samostatný prvok - dvere s ich jedinečným identifikátorom a atribútmi a spojenie s miestnosťou sa uskutočňuje prostredníctvom parametra "ID miestnosti".



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Obr. 3.2-7 Informácie o troch prvkoch kategórie "Dvere" projektu v tabuľkovej štruktúrovanej forme.

Vo voľne štruktúrovaných formátoch, ako je JSON alebo XML, sú údaje uložené v hierarchickej alebo vnorenej forme, pričom prvky môžu obsahovať ďalšie objekty a ich štruktúra sa môže meniť. To umožňuje modelovať zložité vzťahy medzi prvkami. Podobné informácie o dverách v projekte, ktoré boli zaznamenané v štruktúrovanej forme (obrázok 3.2-7), sú reprezentované vo voľne štruktúrovanom formáte (JSON) takým spôsobom (obrázok 3.2-8), že sa stávajú vnorenými objektmi v rámci miestností (Rooms - ID), čo logicky odráža hierarchiu.

```

1  {
2    "Rooms": [
3      {
4        "ID": 101,
5        "Doors": [
6          {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7          {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8        ]
9      },
10     {
11       "ID": 102,
12       "Doors": [
13         {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14       ]
15     }
16   ]
17 }
18

```

The screenshot shows a JSON file named 'doors.json' with two rooms. Room 101 contains two doors (ID 1 and ID 2) with specific dimensions and fireproofing status. Room 102 contains one door (ID 3) with its dimensions and fireproofing status.

Obr. 3.2-8 Informácie o prvkoch kategórie "Dvere" projektu vo formáte JSON.

V modeli grafu sú údaje reprezentované ako uzly (vrcholy) a prepojenia (hrany) medzi nimi. To umožňuje vizualizovať zložité vzťahy medzi objektmi a ich atribútmi. V prípade údajov o dverách a miestnostiach v projekte je reprezentácia grafu nasledovná:

- **Uzly (nodes)** predstavujú hlavné entity: miestnosti (Room 101, Room 102) a dvere (ID1001, ID1002, ID1003).
- **Rebrá (odkazy)** ukazujú vzťahy medzi týmito entitami, napr. príslušnosť dverí k určitej miestnosti
- **Atribúty** sú mapované na uzly a obsahujú vlastnosti entity (výška, šírka, požiarna odolnosť dverí).

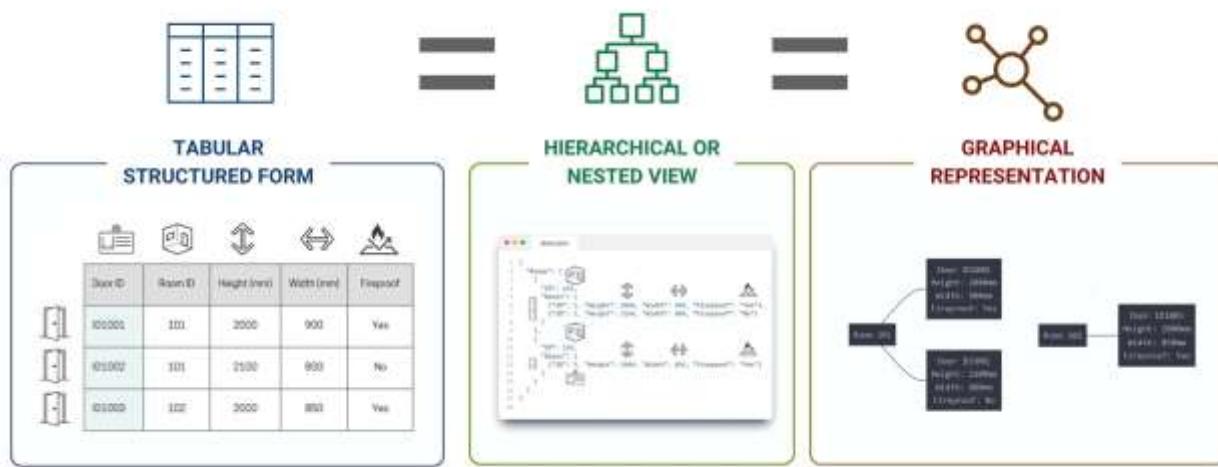


Obr. 3.2-9 Informácie o entite projektových dverí v grafickom zobrazení.

V grafovom dátovom modeli popisu dverí sú každá miestnosť a každé dvere samostatnými uzlami. Dvere sú prepojené s miestnosťami prostredníctvom hrán, ktoré označujú, že dvere patria do konkrétnej miestnosti. Atribúty dverí (výška, šírka, požiarna odolnosť) sú uložené ako vlastnosti príslušných uzlov. Podrobnejšie informácie o grafových formátoch a o tom, ako sa grafová sémantika objavila v stavebnictve, budú uvedené v kapitole "Vznik sémantiky a ontológie v stavebnictve".

Grafové databázy sú efektívne vtedy, keď sú dôležité skôr vzťahy medzi údajmi než samotné údaje, napríklad v odporúčacích systémoch, smerovacích systémoch alebo pri modelovaní zložitých vzťahov v projektoch správy zariadení. Grafový formát zjednodušuje vytváranie nových vzťahov tým, že umožňuje pridať do grafu nové typy údajov bez toho, aby sa zmenila štruktúra úložiska. V porovnaní s relačnými tabuľkami a štruktúrovanými formátm však v grafe nie je žiadna dodatočná dátová konektivita - prenosom dvojrozmerných databázových údajov do grafu sa nezvyšuje počet vzťahov a neposkytujú sa nové informácie.

Forma a schéma údajov by mala byť prispôsobená konkrétnemu prípadu použitia a úlohám, ktoré sa majú vykonať. Na efektívnu prácu v podnikových procesoch je dôležité používať také nástroje a také dátové modely, ktoré pomáhajú čo najrýchlejšie a najjednoduchšie získať výsledky.



Obrázok 3.2-10 Rovnaké informácie o prvkoch projektu môžu byť uložené v rôznych formátoch pomocou rôznych dátových modelov.

Väčšina veľkých spoločností dnes čeli problému nadmernej zložitosti údajov. Každá zo stoviek alebo tisícok aplikácií používa vlastný dátový model, čo vytvára nadmernú zložitosť - jednotlivý model je často desiatky krát zložitejší, ako je potrebné, a súhrn všetkých modelov je tisíckrát zložitejší. Táto nadmerná zložitosť výrazne stáže prácu vývojárom aj koncovým používateľom.

Takáto zložitosť spôsobuje vážne obmedzenia pri vývoji a údržbe systémov spoločnosti. Každý nový prvk v modeli si vyžaduje dodatočný kód, implementáciu novej logiky, dôkladné testovanie a

prispôsobenie existujúcim riešeniam. To všetko zvyšuje náklady a spomaľuje prácu tímu automatizácie v spoločnosti, čím sa aj jednoduché úlohy menia na nákladné a časovo náročné procesy.

Komplexnosť ovplyvňuje všetky úrovne dátovej architektúry. V relačných databázach sa prejavuje v rastúcom počte tabuľiek a stĺpcov, ktoré sú často nadbytočné. V objektovo orientovaných systémoch sa zložitosť zvyšuje množstvom tried a vzájomne súvisiacich vlastností. Vo formátoch, ako je XML alebo JSON, sa zložitosť prejavuje zmätočnými vnorenými štruktúrami, jedinečnými kľúčmi a nekonzistentnými schémami.

Prílišná zložitosť dátových modelov spôsobuje, že systémy sú nielen menej efektívne, ale aj ľažko zrozumiteľné pre koncových používateľov a v budúcnosti sa budú používať veľké jazykové modely a agenti LLM. Práve problém pochopenia a zložosti dátových modelov a spracovania údajov vyvoláva otázku: ako urobiť údaje dostatočne jednoduché na používanie, aby skutočne rýchlo začali byť užitočné.

Aj keď sú dátové modely zvolené rozumne, ich užitočnosť sa dramaticky znižuje, ak je prístup k údajom obmedzený. Proprietárne formáty a uzavreté platformy bránia integráciu, komplikujú automatizáciu a odnímajú kontrolu nad vlastnými informáciami, čím vytvárajú nielen silo nových údajov, ale aj uzamknuté silo, ku ktorému je možné pristupovať len na základe povolenia dodávateľa. Na pochopenie rozsahu problému je dôležité zvážiť, ako presne uzavreté systémy ovplyvňujú digitálne procesy v stavebnictve.

## **Proprietárne formáty a ich vplyv na digitálne procesy**

Jednou z kľúčových výziev, ktorým stavebné spoločnosti čelia počas digitalizácie, je obmedzený prístup k údajom. To stáhuje integráciu systémov, znižuje kvalitu informácií a stáhuje organizáciu efektívnych procesov. Príčinou týchto ľažkostí je často používanie proprietárnych formátov a uzavretých softvérových riešení.

Bohužiaľ, doteraz mnohé programy používané v stavebnictve umožňujú používateľovi ukladať údaje len v proprietárnych formátoch alebo v cloudových úložiskách, ku ktorým možno pristupovať len prostredníctvom prísne obmedzených rozhraní. A nezriedka sú tieto riešenia postavené v závislosti od ešte uzavretejších systémov od väčších dodávateľov. V dôsledku toho sú aj tí vývojári, ktorí by chceli ponúkať otvorenejšie architektúry, nútení dodržiavať pravidlá diktované veľkými dodávateľmi.

Hoci moderné systémy správy stavebných údajov čoraz viac podporujú otvorené formáty a štandardy (obr. 3.1-5), databázy založené na CAD (BIM) a súvisiace systémy ERP a CAFM zostávajú izolovanými proprietárnymi "ostrovmi" v digitálnom prostredí odvetvia (obr. 3.2-11).

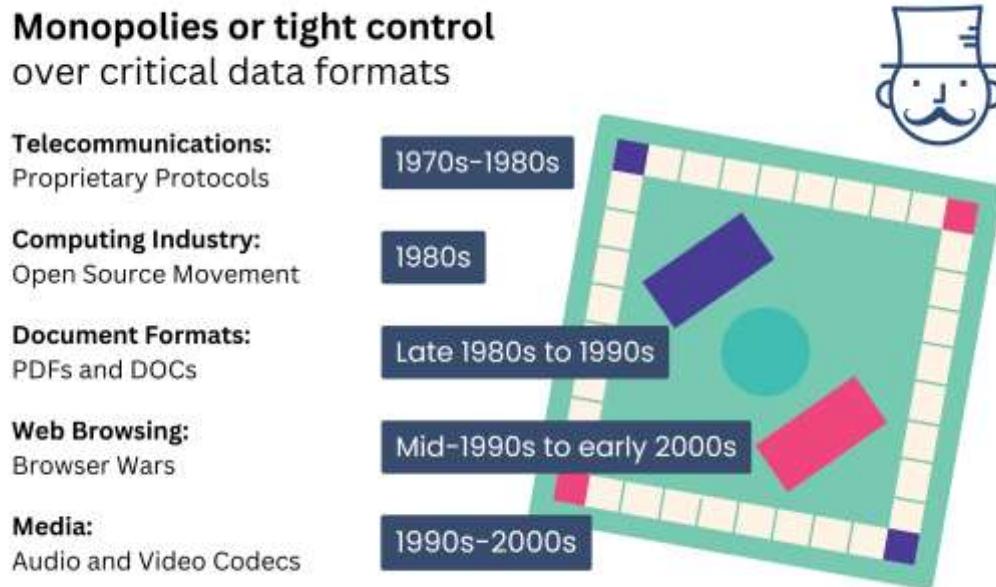


Obrázok 3.2-11 Uzavretá a proprietárna povaha údajov vytvára prekážky pre integráciu a prístup k údajom.

Uzavreté a monopolizované formáty a protokoly nie sú problémom len pre stavebníctvo. V mnohých odvetviach hospodárstva sa boj proti uzavretým štandardom a obmedzenému prístupu k údajom začal spomalnením inovácií (obrázok 3.2-12), existenciou umelých prekážok vstupu nových subjektov na trh a prehlbovaním závislosti od veľkých dodávateľov. S rýchlym rastom významu údajov orgány pre hospodársku súťaž jednoducho nemajú čas reagovať na výzvy, ktoré predstavujú nové digitálne trhy, a v dôsledku toho sa uzavreté formáty a obmedzený prístup k údajom v podstate stávajú digitálnymi "hranicami", ktoré obmedzujú tok informácií a rast [63].

*Ak stroje vyrobia všetko, čo potrebujeme, naša situácia bude závisieť od toho, ako sa tieto tovary budú distribuovať. Každý si bude môcť užívať život v blahobye len vtedy, ak sa bude bohatstvo vyrobené strojmi rozdeľovať. Alebo väčšina ľudí skončí v krajnej chudobe, ak majitelia áut budú môcť úspešne lobovať proti prerozdeľovaniu bohatstva. Zatiaľ sa zdá, že veci sa vyvíjajú druhým smerom, pričom technológie vedú k stále väčšej nerovnosti [72].*

- Stephen Hawking, astrofyzik, 2015



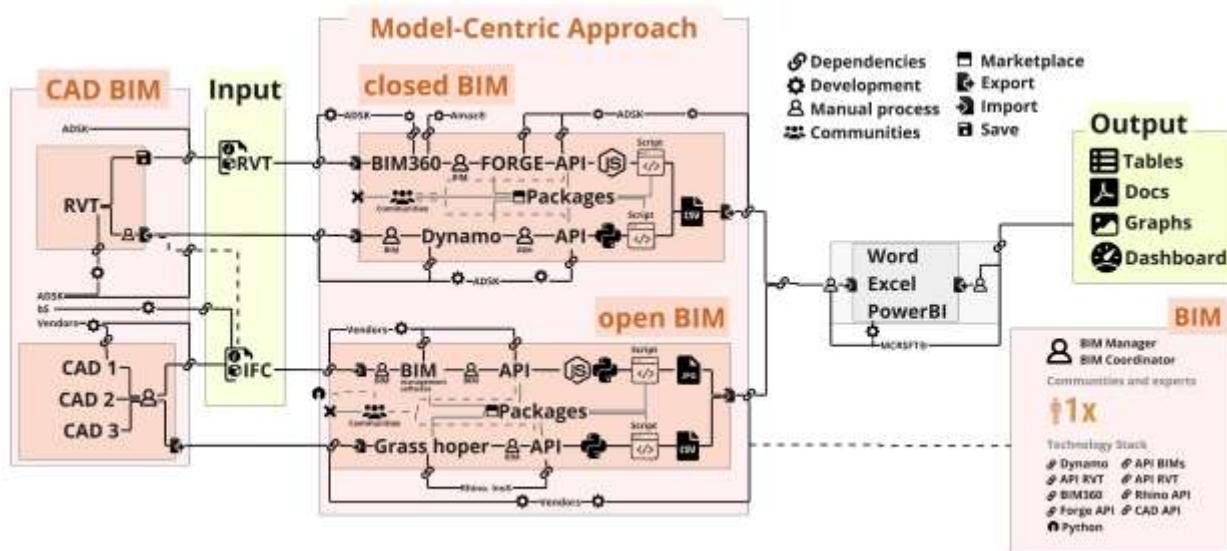
Obr. 3.2-12 Monopolné vlastníctvo kľúčových dátových formátov a protokolov nie je výlučným problémom stavebníctva.

V dôsledku uzavretého prístupu k databázam programy, manažéri údajov, dátoví analytici, IT špecialisti a vývojári vytvárajúci aplikácie na prístup k údajom, ich spracovanie a automatizáciu v stavebníctve dnes čelia početným závislostiam od dodávateľov softvéru (obr. 3.2-13). Tieto závislosti v podobe ďalších prístupových vrstiev si vyžadujú vytváranie riešení so špecializovanými API -prípojkami a špeciálnymi nástrojmi a softvérom.

Rozhranie API (Application Programming Interface) je formalizované rozhranie, prostredníctvom ktorého môže jeden program komunikovať s druhým a vymieňať si údaje a funkcie bez toho, aby musel pristupovať k zdrojovému kódu. Rozhranie API opisuje, aké požiadavky môže externý systém zadávať, v akom formáte by mali byť a aké odpovede dostane. Je to štandardizovaná "zmluva" medzi softvérovými modulmi.

Veľké množstvo závislostí na uzavretých riešeniacach spôsobuje, že celá architektúra kódu a logika obchodných procesov v podniku sa stáva "špagetovou architektúrou" nástrojov závislých od politiky dodávateľa softvéru, ktorý poskytuje kvalitný prístup k údajom.

Závislosť na uzavretých riešeniacach a platformách viedie nielen k strate flexibility, ale aj k reálnym obchodným rizikám. Zmena licenčných podmienok, uzavretie prístupu k údajom, zmena formátov alebo štruktúry API - to všetko môže zablokovať kritické procesy. Zrazu sa ukáže, že aktualizácia jednej tabuľky si vyžaduje prepracovanie celého bloku integrácií a konektorov (obr. 3.2-13) a akákoľvek rozsiahla aktualizácia softvéru alebo jeho dodávateľa API sa stáva potenciálou hrozbou pre stabilitu celého podnikového systému.



Obr. 3.2-13 Príklad veľkého počtu závislostí pri spracovaní CAD -údajov vytvára prekážky pri integrácii údajov v ekosystéme stavebných spoločností.

Vývojári a systémoví architekti sú v takýchto podmienkach nútení pracovať nie na očakávaní, ale na prežití. Namiesto implementácie nových riešení sa prispôsobujú. Namiesto vývoja sa snažia zachovať kompatibilitu. Namiesto automatizácie a zrýchľovania procesov trávia čas štúdiom ďalších uzavretých rozhraní, dokumentácie API a nekonečným prestavovaním kódu.

Práca s uzavretými formátmi a systémami nie je len technickou výzvou - je to strategické obmedzenie. Napriek zjavným možnostiam, ktoré ponúka moderná automatizácia, umelá inteligencia, LLM a prediktívna analýza, mnohé spoločnosti nedokážu naplno využiť ich potenciál. A prekážky, ktoré kladú proprietárne formáty (obrázok 3.2-13), bránia podnikom v prístupe k vlastným údajom. To je možno irónia digitálnej transformácie v stavebníctve.

Transparentnosť údajov a otvorené systémy nie sú luxusom, ale nevyhnutnou podmienkou rýchlosťi a efektívnosti. Bez otvorenosti sú obchodné procesy plné zbytočnej byrokracie, viacvrstvových schvaľovacích reťazcov a rastúcej závislosti od princípu HiPPO - rozhodovanie na základe názoru najlepšie platenej osoby.

Napriek tomu sa na obzore črtá zmena paradigmy. Napriek dominancii proprietárnych riešení si čoraz viac spoločnosti uvedomuje obmedzenia architektúr inšpirovaných štvrtou priemyselnou revolúciou. V súčasnosti sa vektor posúva smerom k princípm piatej revolúcie, kde sú v centre pozornosti údaje ako strategické aktívum, otvorené rozhrania (API) a skutočná interoperabilita medzi systémami.

Tento prechod znamená posun od uzavretých ekosystémov k flexibilným, modulárnym digitálnym architektúram, kde sú kľúčové otvorené formáty, štandardy a transparentná výmena údajov.

## Otvorené formáty menia prístup k digitalizácii

Stavebníctvo bolo jedným z posledných odvetví, ktoré riešilo problém uzavretých a chránených údajov. Na rozdiel od iných odvetví hospodárstva sa tu digitalizácia rozvíjala pomaly. Medzi dôvody patrí tradičný konzervatívny charakter odvetvia, prevaha rozdielnych lokálnych riešení a hlboko zakorenená povaha správy dokumentov v papierovej podobe. Klúčové procesy v stavebníctve sa po desaťročia spoliehali na fyzické výkresy, telefonáty a nesynchronizované databázy. V tomto kontexte boli uzavreté formáty dlho vnímané skôr ako norma než ako prekážka.

Skúsenosti z iných priemyselných odvetví ukazujú, že odstránenie prekážok pre uzavreté údaje vedie k prudkému nárastu inovácií, zrýchleniu vývoja a zvýšeniu konkurencie [73]. Vo vede výmena otvorených údajov umožňuje urýchliť objavy a rozvíjať medzinárodnú spoluprácu. V medicíne môže zlepšiť účinnosť diagnostiky a liečby. V softvérovom inžinierstve - vytvárať ekosystémy spoluvorby a rýchleho zlepšovania produktov.

Podľa správy spoločnosti McKinsey "Open Data: Odomknite inovácie a produktivitu pomocou toku informácií" 2013. [74] majú otvorené údaje potenciál uvoľniť 3 až 5 biliónov USD ročne v siedmich klúčových odvetviach vrátane stavebníctva, dopravy, zdravotníctva a energetiky. Podľa tej istej štúdie decentralizované dátové ekosystémy umožňujú veľkým stavebným spoločnostiam a dodávateľom znížiť náklady na vývoj a údržbu softvéru, čo urýchluje zavádzanie digitálnych technológií.

Prechod na otvorené architektúry, ktorý sa už dávno začal v iných odvetviach hospodárstva, sa postupne šíri aj do stavebníctva. Veľké spoločnosti a verejný klienti, a najmä finančné organizácie, ktoré kontrolujú investície do stavebných projektov, čoraz viac požadujú používanie otvorených údajov a prístup k zdrojovému kódovi výpočtov, kalkulácií a aplikácií. Od vývojárov sa už neočakáva len vytvorenie digitálnych riešení a zobrazenie konečných čísel projektu - očakáva sa, že budú transparentné, reprodukovateľné a nezávislé od dodávateľov aplikácií tretích strán.

Používanie riešení s otvoreným zdrojovým kódom poskytuje zákazníkovi istotu, že aj keď externí vývojári prestanú spolupracovať alebo opustia projekt, neovplyvní to schopnosť ďalej vyvíjať nástroje a systémy. Jednou z hlavných výhod otvorených údajov je ich schopnosť odstrániť závislosť vývojárov aplikácií od konkrétnych platform na prístup k údajom.

Ak sa spoločnosť nemôže úplne vzdať proprietárnych riešení, možným kompromisom je použitie techník reverzného inžinierstva. Tieto právne a technicky správne metódy umožňujú konvertovať uzavreté formáty na prístupejšie, štruktúrované a vhodné na integráciu. To je dôležité najmä v prípade potreby pripojenia k starším systémom alebo migrácie informácií z jedného softvérového prostredia do druhého.

Jedným z najjasnejších príkladov v histórii prechodu na otvorené formáty a uplatňovania reverzného inžinierstva (legálne nabúravanie proprietárnych systémov) v stavebnictve je história boja za otvorenie formátu DWG, ktorý sa široko používa v systémoch počítačom podporovaného projektovania (CAD). V roku 1998 v reakcii na monopol jedného dodávateľa softvéru vytvorilo ďalších 15 dodávateľov CAD novú alianciu s názvom "Open DWG" s cieľom poskytnúť vývojárom bezplatné a nezávislé nástroje na prácu s formátom DWG (de facto štandard na prenos výkresov) bez potreby proprietárneho softvéru alebo uzavretých API. Táto udalosť bola zlomovým bodom, ktorý umožnil desaťtisícom spoločností získať bezplatný prístup k uzatvorenému formátu populárneho riešenia CAD od konca 80. rokov až dodnes a vytvoriť kompatibilné riešenia, ktoré podporili konkurenciu na trhu CAD [75]. Dnes sa SDK "Open DWG", ktorý bol prvýkrát vytvorený už v roku 1996, používa takmer vo všetkých riešeniacach, v ktorých je možné importovať, upravovať a exportovať formát DWG mimo oficiálnej aplikácie tvorca formátu DWG.

Podobné zmeny si vynucujú aj ďalší technologickí giganti. Spoločnosť Microsoft, kedysi symbol proprietárneho prístupu, otvorila zdrojový kód .NET Framework, začala používať Linux v infraštruktúre cloudovej služby Azure a získala GitHub, aby posilnila svoju pozíciu v komunité open source. [76]. Spoločnosť Meta (predtým Facebook) zverejnila modely AI s otvoreným zdrojovým kódom, ako napríklad sériu Llama, s cieľom podporiť inovácie a spoluprácu pri vývoji agentov AI. Generálny riaditeľ Mark Zuckerberg predpokladá, že platformy s otvoreným zdrojovým kódom budú v nasledujúcom desaťročí viestť technologický pokrok [77].

Open Source je model vývoja a distribúcie softvéru, v ktorom je zdrojový kód otvorený na voľné používanie, štúdium, modifikáciu a distribúciu.

Otvorené údaje a riešenia s otvoreným zdrojovým kódom sa stávajú nielen trendom, ale aj základom digitálnej udržateľnosti. Poskytujú spoločnostiam flexibilitu, odolnosť, kontrolu nad vlastnými rozhodnutiami a možnosť škálovať digitálne procesy bez závislosti od politík dodávateľov. A čo je rovnako dôležité, vracajú podnikom kontrolu nad najcennejším zdrojom 21. storočia - ich údajmi.

## Zmena paradigmy: Open Source ako koniec éry dominancie dodávateľov softvéru

Stavebnictvo prechádza zmenou, ktorá sa nedá speňažiť bežným spôsobom. Koncepcia prístupu založeného na údajoch, orientovaného na dáta a využívanie nástrojov s otvoreným zdrojovým kódom vedie k prehodnoteniu pravidiel hry, na ktorých stoja softvéroví giganti na trhu.

Na rozdiel od predchádzajúcich technologických transformácií tento prechod nebude aktívne podporovaný dodávateľmi. Zmena paradigmy ohrozenie ich tradičné obchodné modely založené na licenciách, predplatnom a konzultáciách. Nová realita nezahrňa hotový produkt alebo platené predplatné - vyžaduje si zmenu procesov a myšlienia.

Na riadenie a vývoj riešení dátových centier založených na otvorených technológiách budú musieť spoločnosti prehodnotiť interné procesy. Špecialisti z rôznych oddelení budú musieť nielen spolupracovať, ale aj prehodnotiť spôsob svojej spolupráce.

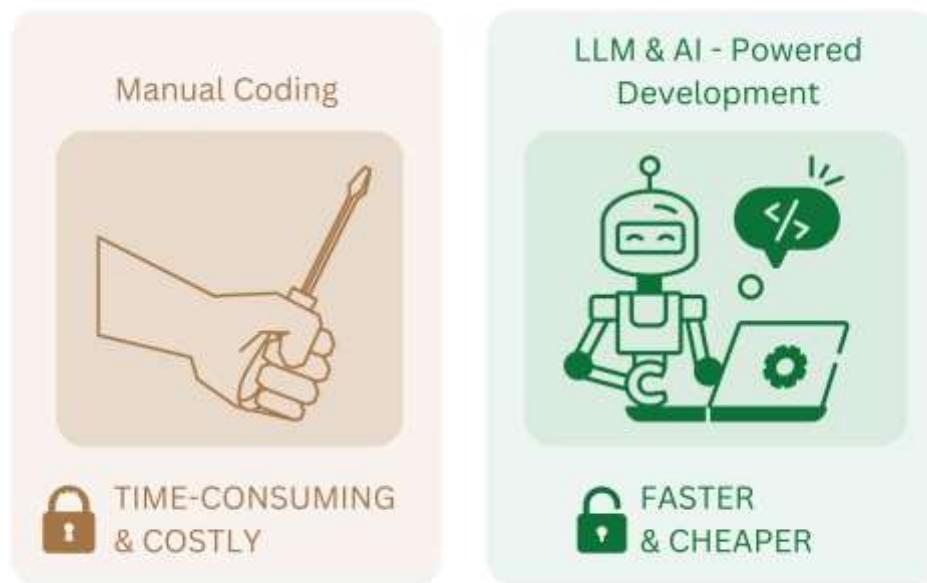
Nová paradigma predpokladá využívanie otvorených údajov a riešení s otvoreným zdrojovým kódom, pričom osobitnú úlohu pri tvorbe softvérového kódu budú zohrávať nástroje založené na umelej inteligencii a veľkých jazykových modeloch (LLM), a nie programátori. Už v polovici roka 2024 sa viac ako 25% nového kódu v spoločnosti Google vytvorí pomocou umelej inteligencie [78]. V budúcnosti bude kódovanie pomocou LLM vykonávať 80% práce len za 20% času (obrázok 3.2-14).

Podľa štúdie McKinsey 2020 [79] GPU čoraz viac nahradzajú CPU v analytike vďaka svojmu vysokému výkonu a podpore moderných nástrojov s otvoreným zdrojovým kódom. To umožňuje spoločnostiam zrýchliť spracovanie údajov bez výrazných investícií do drahého softvéru alebo najímania nedostatkových špecialistov.

Popredné poradenské spoločnosti, ako napríklad McKinsey, PwC a Deloitte, zdôrazňujú rastúci význam otvorených štandardov a aplikácií s otvoreným zdrojovým kódom v rôznych odvetviach.

Podľa správy PwC Open Source Monitor 2019 [80] 69% spoločností so 100 a viac zamestnancami vedome používa riešenia s otvoreným zdrojovým kódom. OSS sa aktívne používa najmä vo veľkých spoločnostiach: 71% spoločností s 200 až 499 zamestnancami, 78% v kategórii 500 až 1999 zamestnancov a až 86% medzi spoločnosťami s viac ako 2000 zamestnancami. Podľa správy Synopsys OSSRA 2023 96% analyzovaných kódových báz obsahovalo komponenty s otvoreným zdrojovým kódom [81].

Úlohou vývojára v budúcnosti nebude ručne písanie kódu, ale navrhovať dátové modely, architektúry tokov a spravovať agentov umelej inteligencie, ktorí na požiadanie vytvoria správne výpočty. Používateľské rozhrania budú minimalistické a interakcia bude založená na dialógu. Klasické programovanie ustúpi návrhu na vysokej úrovni a orchestrácii digitálnych riešení (obrázok 3.2-14). Súčasné trendy - ako napríklad platformy s nízkym kódom (obr. 7.4-6) a ekosystémy s podporou LLM (obr. 7.4-4) - výrazne znížia náklady na vývoj a údržbu IT systémov.



Obr. 3.2-14 Zatiaľ čo dnes aplikácie vytvárajú programátori manuálne, v budúcnosti bude značná časť kódu generovaná pomocou riešení založených na umelej inteligencii a LLM.

Tento prechod bude odlišný od predchádzajúcich a veľkí dodávatelia softvéru pravdepodobne nebudú katalyzátormi.

V štúdii Harvard Business School "The Value of Open Source Software" 2024 [40] sa celková hodnota open source softvéru odhaduje z dvoch hľadísk. Na jednej strane, ak by sme vypočítali, kolko by bolo potrebné na vytvorenie všetkých existujúcich riešení s otvoreným zdrojovým kódom od začiatku, suma by bola približne 4,15 miliardy USD. Na druhej strane, ak si predstavíme, že každá spoločnosť si sama vyvíja vlastné analógie Open Source riešení (čo sa deje všade), bez toho, aby mala prístup k existujúcim nástrojom, potom by celkové náklady na podnikanie dosiahli kolosalných 8,8 bilióna dolárov - to je cena dopytu.

Nie je ľahké uhádnuť, že žiaden veľký výrobca softvéru nemá záujem na zmenšenie softvérového trhu s potenciálnou hodnotou 8,8 bilióna dolárov na iba 4,15 miliardy dolárov. To by znamenalo znížiť objem dopytu viac ako 2 000-krát. Takáto transformácia je pre dodávateľov, ktorých obchodné modely sú postavené na dlhoročnom udržiavaní závislosti zákazníkov od uzavretých riešení, jednoducho nerentabilná. Takže spoločnosti, ktoré očakávajú, že im niekto ponúkne pohodlné a otvorené riešenie na kľúč, môžu byť sklamané - títo dodávatelia sa jednoducho neobjavia.

Prechod na otvorenú digitálnu architektúru neznamená stratu pracovných miest alebo príjmov. Naopak, vytvára podmienky pre flexibilné a prispôsobivé obchodné modely, ktoré môžu nakoniec vytlačiť tradičný trh s licenciami a krabicovým softvérom.

Namiesto predaja licencií - služby, namiesto uzavretých formátov - otvorené platformy, namiesto zá-

vislosti od dodávateľa - nezávislosť a možnosť vytvárať riešenia pre skutočné potreby. Tí, ktorí doteraz nástroje len používali, sa budú môcť stať ich spoluautormi. A tí, ktorí dokážu pracovať s údajmi, modelmi, scenármi a logikou, sa ocitnú v centre novej digitálnej ekonomiky priemyslu. O týchto zmenách a o tom, aké nové úlohy, obchodné modely a formáty spolupráce vznikajú okolo otvorených údajov, budeme viac hovoriť v záverečnej, desiatej časti knihy.

Riešenia založené na otvorených údajoch a otvorenom kóde umožnia spoločnostiam sústrediť sa na efektívnosť obchodných procesov, a nie na boj so zastaranými API a integráciu uzavretých systémov. Vedomý prechod na otvorenú architektúru môže výrazne zvýšiť produktivitu a znížiť závislosť od dodávateľov.

Prechod do novej reality neznamená len zmenu prístupov k vývoju softvéru, ale aj prehodnotenie samotného princípu práce s údajmi. V centre tejto transformácie nie je kód, ale informácie: ich štruktúra, dostupnosť a interpretovateľnosť. Tu sa do popredia dostávajú otvorené a štruktúrované údaje, ktoré sa stávajú neoddeliteľnou súčasťou novej digitálnej architektúry.

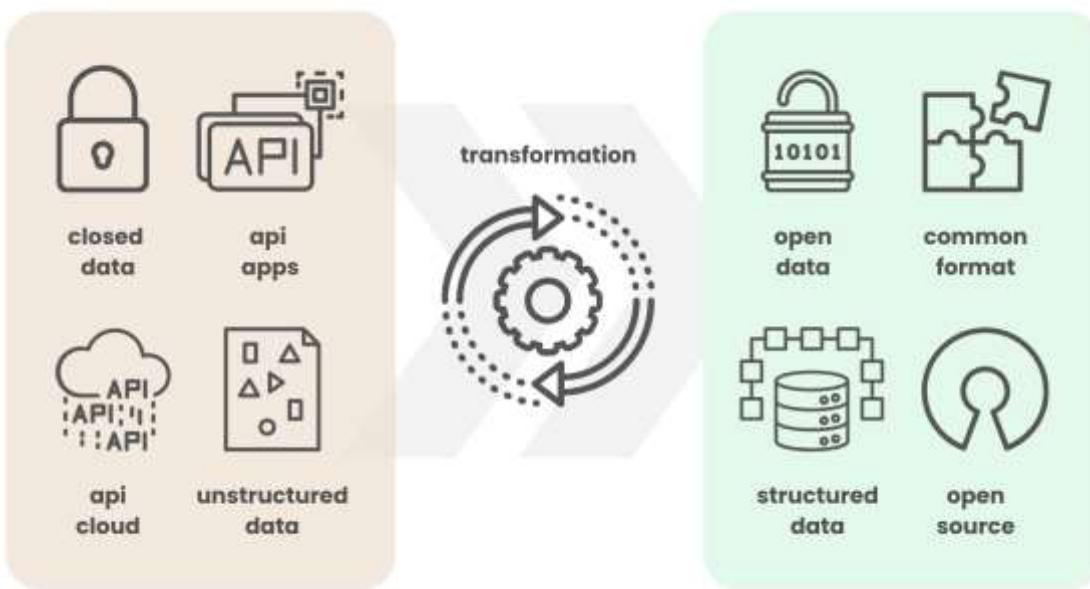
## Štruktúrované otvorené údaje: základ digitálnej transformácie

Zatiaľ čo v minulých desaťročiach bola udržateľnosť podnikov do veľkej miery podmienená výberom softvérových riešení a závislosťou od konkrétnych dodávateľov, v dnešnej digitálnej ekonomike je klúčovým faktorom kvalita údajov a schopnosť s nimi efektívne pracovať. Otvorený zdrojový kód je dôležitou súčasťou novej technologickej paradigmy, ale jeho potenciál sa skutočne uvoľní len vtedy, keď sú údaje zrozumiteľné, usporiadane a strojovo čitateľné. Spomedzi všetkých typov dátových modelov sa štruktúrované otvorené údaje stávajú základným kameňom udržateľnej digitálnej transformácie.

Hlavnou výhodou štruktúrovaných otvorených údajov je jednoznačná interpretácia a možnosť automatizovaného spracovania. To umožňuje výrazné zvýšenie efektívnosti na úrovni jednotlivých operácií aj v rámci celej organizácie.

Podľa správy spoločnosti Deloitte "Proces prenosu údajov pri transformácii podnikov" [82] je spolupráca s IT oddelením pri riadení prenosu štruktúrovaných údajov veľmi dôležitá. Podľa správy vlády Spojeného kráľovstva Data Analytics and AI in Government Project Delivery (2024) [83] je odstránenie prekážok pri zdieľaní údajov medzi rôznymi projektmi a organizáciami klúčom k zvýšeniu efektívnosti riadenia projektov. V dokumente sa zdôrazňuje, že štandardizácia formátov údajov a zavedenie zásad otvorených údajov môže zabrániť duplicitné informácií, minimalizovať časové straty a zlepšiť presnosť prognóz.

Pre stavebníctvo, kde tradične prevláda vysoká miera roztriedenosťi a rôznorodosťi formátov, zohráva proces štruktúrovaného zjednotenia a štruktúrované otvorené údaje klúčovú úlohu pri formovaní koherentných a riadených procesov (obr. 4.1-14). Umožňujú účastníkom projektu zamerať sa skôr na zvýšenie produktivity než na riešenie technických problémov súvisiacich s nekompatibilitou uzavretých platforem, dátových modelov a formátov.



Obr. 3.2-15 Otvorené štruktúrované údaje znižujú závislosť od softvérových riešení a platformiem a urýchľujú inovácie.

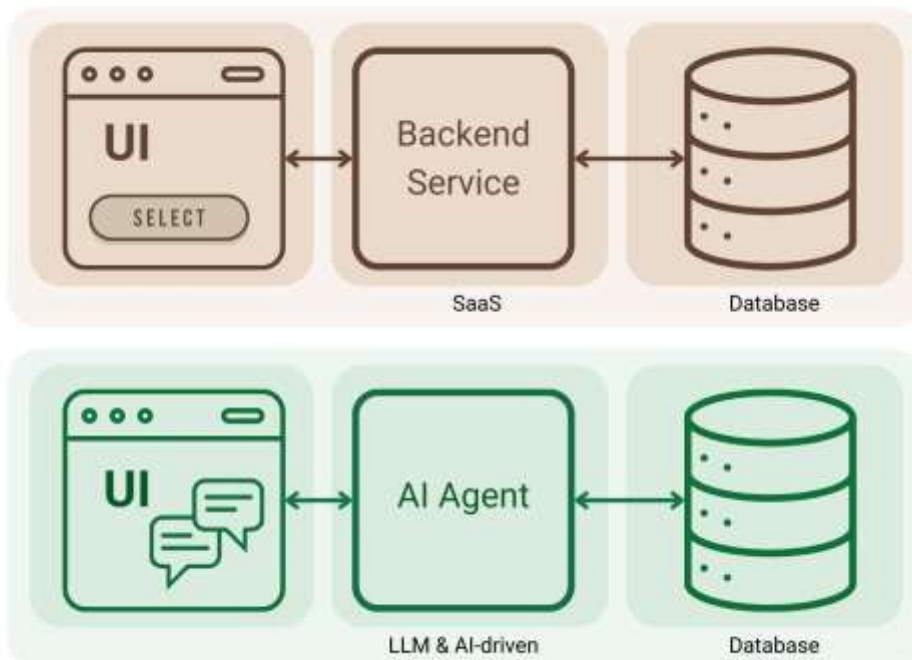
Moderné technologické nástroje, ktorým sa budeme podrobne venovať neskôr v knihe, umožňujú nielen zhromažďovať informácie, ale aj automaticky ich čísiť: odstraňovať duplicitu, opravovať chyby a normalizovať hodnoty. To znamená, že analytici a inžinieri nepracujú s rôznorodými dokumentmi, ale s usporiadanou bázou znalostí vhodnou na analýzu, automatizáciu a rozhodovanie.

*Urobte to čo najjednoduchšie, ale nie jednoduchšie.*

- Albert Einstein, teoretický fyzik (autorstvo citátu je sporné [84])

Dnes sa väčšina používateľských rozhranií na prácu s údajmi dá vytvoriť automaticky - bez potreby ručného písania kódu pre každý obchodný prípad. To si vyžaduje vrstvu infraštruktúry, ktorá rozumie štruktúre údajov, modelu a logike bez dodatočných inštrukcií (obrázok 4.1-15). Tento prístup umožňujú práve štruktúrované údaje: formuláre, tabuľky, filtre a pohľady sa môžu automaticky generovať s minimálnym programátorským úsilím.

Najdôležitejšie rozhrania, ktoré sú pre používateľa kritické, si môžu vyžadovať manuálne zdokonalenie. Vo väčšine prípadov - a to je 50 až 90% pracovných scenárov - však postačuje automatické generovanie aplikácií a výpočtov bez použitia špeciálnych aplikácií na tento účel (obr. 3.2-16), čo výrazne znižuje náklady na vývoj a údržbu, znižuje chybovosť a urýchľuje implementáciu digitálnych riešení.



Obr. 3.2-16 Architektonické modely na prácu s údajmi: tradičná aplikáčna architektúra a model orientovaný na umelú inteligenciu s LLM.

Ďalším krokom v digitálnej evolúcii je prechod od architektúr postavených na jednotlivých aplikáciách k inteligentne riadeným systémom založeným na jazykových modeloch (LLM). V takejto architektúre sa štruktúrované údaje stávajú nielen predmetom ukladania, ale aj základom pre interakciu s nástrojmi umelej inteligencie, ktoré sú schopné analyzovať, interpretovať a odporúčať činnosti na základe kontextu.

V nasledujúcich kapitolách sa pozrieme na reálne príklady implementácie architektúry založenej na otvorených štruktúrovaných údajoch a ukážeme si, ako sa jazykové modely používajú na automatickú interpretáciu, validáciu a spracovanie údajov. Tieto praktické prípady vám pomôžu lepšie pochopiť, ako nová digitálna logika funguje v praxi - a aké výhody prináša spoločnostiam, ktoré sú pripravené na transformáciu.



## KAPITOLA 3.3.

### **LLM A ICH ÚLOHA PRI SPRACOVANÍ ÚDAJOV A OBCHODNÝCH PROCESOCH**

#### **Chatovacie miestnosti LLM: ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok na automatizáciu dátových procesov.**

Vznik veľkých jazykových modelov (Large Language Models - LLM) bol prirodzeným pokračovaním hnutia za štruktúrované otvorené údaje a filozofie otvoreného softvéru. Keď sa údaje stanú organizovanými, prístupnými a strojovo čitateľnými, ďalším krokom je nástroj, ktorý dokáže s týmito informáciami pracovať bez potreby písat zložitý kód alebo mať špecializované technické znalosti.

LLM sú priamym produktom otvorenosti: rozsiahle otvorené súbory údajov, publikácie a hnutie Open Source. Bez otvorených vedeckých článkov, verejne dostupných textových údajov a kultúry spoločného vývoja by neexistoval ChatGPT ani iné LLM. LLM je v istom zmysle "destilátom" nahromadených digitálnych vedomostí ľudstva, ktoré sa zhromažďujú a vzdelávajú prostredníctvom princípov otvorenosti.

Moderné veľké jazykové modely (LLM - Large Language Models), ako napríklad ChatGPT ® (OpenAI), LlaMa ™ (Meta AI), Mistral DeepSeek™, Grok ™ (xAI), Claude ™ (Anthropic), QWEN™ poskytujú používateľom možnosť formulovať dotazy na údaje v prirodzenom jazyku. Tým sa práca s informáciami sprístupňuje nielen vývojom, ale aj analytikom, inžinierom, plánovačom, manažérom a ďalším odborníkom, ktorým bolo programovanie predtým vzdialené

**LLM** (Large Language Model) je umelá inteligencia, ktorá je vyškolená na porozumenie a generovanie textu na základe obrovského množstva údajov zozbieraných z celého internetu. Dokáže analyzovať kontext, odpovedať na otázky, zapájať sa do dialógov, písat texty a generovať softvérový kód.

Ak sa predtým na vizualizáciu, spracovanie alebo analýzu údajov vyžadovala znalosť špeciálneho programovacieho jazyka: Python, SQL, R alebo Scala, ako aj schopnosť pracovať s knižnicami ako Pandas, Polars alebo DuckDB a mnohými ďalšími, od roku 2023 sa situácia radikálne zmenila. Teraz môže používateľ jednoducho opísať, čo chce získať - a model sám vygeneruje kód, vykoná ho, zobrazí tabuľku alebo graf a vysvetlí výsledok. Prvýkrát po desaťročiach sa vývoj technológií nevydal cestou komplikovania, ale radikálneho zjednodušenia a sprístupnenia.

Tento princíp - "spracovávať údaje pomocou slov (podnetov)". - znamenal novú etapu vo vývoji práce s informáciami, čím sa tvorba riešení dostala na ešte vyššiu úroveň abstrakcie. Tak ako kedysi už nebolo potrebné, aby používateelia rozumeli technickým základom internetu, aby mohli prevádzkovať internetové obchody alebo vytvárať webové stránky pomocou WordPress, Joomla a iných modulárnych systémov s otvoreným zdrojovým kódom (av tor knihy pracuje s takýmito systémami od

roku 2005, vrátane vzdelávacích a inžinierskych online platform). - To následne viedlo k rozmachu digitálneho obsahu a online podnikania - inžinieri, analytici a manažéri dnes môžu automatizovať pracovné postupy bez znalosti programovacích jazykov. Napomáhajú tomu výkonné LLM - bezplatné aj open source, ako napríklad LLaMA, Mistral, Qwen, DeepSeek a ďalšie - ktoré sprístupňujú pokročilé technológie čo najširšiemu publiku.

## **Veľké jazykové modely LLM: ako to funguje**

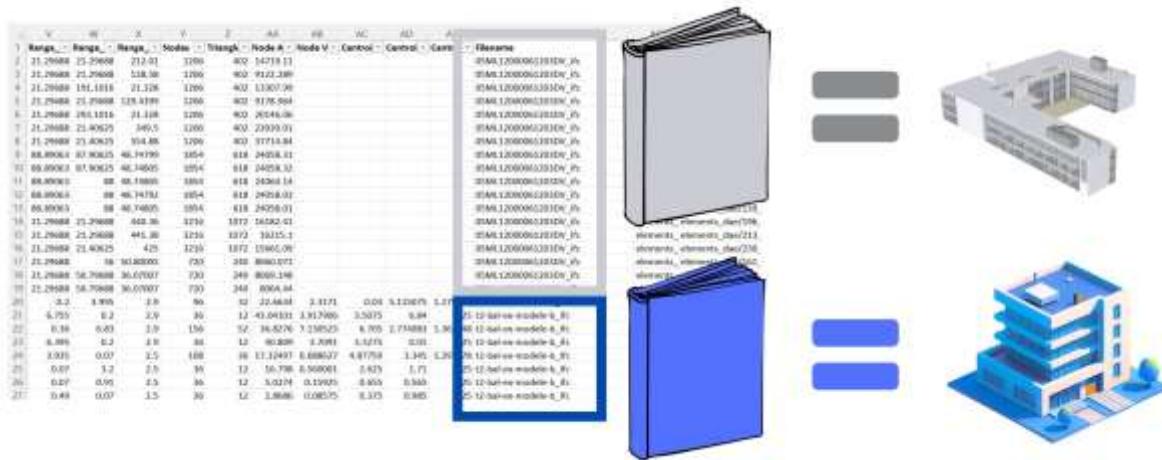
Veľké jazykové modely (ChatGPT, LLaMA, Mistral, Claude, DeepSeek, QWEN, Grok) sú neurónové siete vyškolené na obrovskom množstve textových údajov z internetu, kníh, článkov a iných zdrojov. Ich hlavnou úlohou je porozumieť kontextu ľudskej reči a generovať zmysluplné odpovede.

Moderný LLM je založený na architektúre Transformer, ktorú navrhli výskumníci spoločnosti Google v roku 2017 [85]. Kľúčovým komponentom tejto architektúry je mechanizmus pozornosti, ktorý umožňuje modelu zohľadňovať vzťahy medzi slovami bez ohľadu na ich pozíciu v texte.

Proces učenia LLM je vzdialene podobný spôsobu, akým sa ľudia učia jazyk, len je miliónkrát väčší. Model analyzuje miliardy príkladov slov a výrazov, pričom identifikuje vzory v štruktúre jazyka a v logike sémantických prechodov. Celý text sa rozdelí na tokeny - minimálne sémantické jednotky (slová alebo ich časti), ktoré sa potom transformujú na vektory vo viaceromernom priestore (obr. 8.2-2). Tieto vektorové reprezentácie umožňujú stroju "pochopiť" skryté vzťahy medzi pojмami, namiesto toho, aby jednoducho operoval s textom ako so sekvenciou symbolov.

Veľké jazykové modely nie sú len nástroje na generovanie textu. Dokážu rozpoznať význam, nájsť súvislosti medzi pojмami a pracovať s údajmi, aj keď sú prezentované v rôznych formátoch. Hlavné je, aby boli informácie rozdelené do zrozumiteľných modelov a reprezentované ako tokeny, s ktorými dokáže LLM pracovať.

Rovnaký prístup možno uplatniť aj na stavebné projekty. Ak si predstavíme projekt ako druh textu, kde každá budova, prvk alebo konštrukcia je symbolom, môžeme začať spracovávať takéto informácie podobným spôsobom. Stavebné projekty možno prirovnáť ku knihám, ktoré sú usporiadane do kategórií, kapitol a skupín odsekov pozostávajúcich z minimálnych tokenov - prvkov stavebného projektu (obrázok 3.3-1). Prekladom dátových modelov do štruktúrovaného formátu môžeme štruktúrované údaje previesť aj do vektorových báz (obr. 8.2-2), ktoré sú ideálnym zdrojom pre strojové učenie a technológie, ako je LLM.



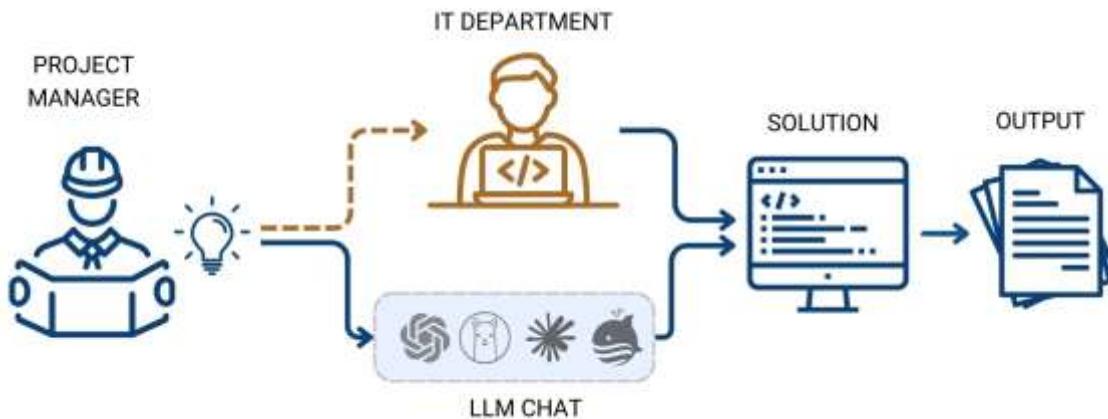
Obr. 3.3-1 Prvok stavebného projektu je ako token v texte: minimálna jednotka, z ktorej sa vytvárajú skupiny (odseky) časťí (kategórií) celého projektu.

Ak je stavebný projekt digitalizovaný a jeho prvky sú reprezentované ako tokeny alebo vektory, je možné k nim pristupovať v prirodzenom jazyku, a nie prostredníctvom prísnych formálnych dotazov. Tu sa prejavuje jedna z kľúčových výhod LLM - schopnosť pochopiť význam dotazu a prepojiť ho s príslušnými údajmi.

Inžinier už nemusí písat SQL -dotaz alebo Python kód, aby získal požadované údaje - môže jednoducho, rozumejúc LLM a štruktúre údajov, formulovať úlohu bežným spôsobom: "Nájdite všetky železobetónové konštrukcie s triedou betónu vyššou ako B30 a vypočítajte ich celkový objem". Model rozpozná význam dotazu, prevedie ho do strojovo čitateľnej podoby, nájde údaje (zoskupí a transformuje) a vráti konečný výsledok.

Dokumenty, tabuľky, modely projektov sa konvertujú na vektorové reprezentácie (vloženie) a ukladajú sa do databázy. Keď používateľ položí otázku, dotaz sa tiež prevedie na vektor a systém nájde najrelevantnejšie údaje. To umožňuje LLM spoliehať sa nielen na vlastné vyškolené znalosti, ale aj na skutočné podnikové údaje, aj keď sa objavili už po skončení trénovalia modelu.

Jednou z najdôležitejších výhod LLM v stavebnictve je možnosť generovať programový kód. Namiesto odovzdania technickej úlohy programátorovi môžu odborníci úlohu opísť v prirodzenom jazyku a model vytvorí potrebný kód, ktorý možno použiť (skopírovaním z chatu) pri tvorbe kódu automatizácie procesov. LLM -modely umožňujú špecialistom bez hlbokých znalostí programovania prispiť k automatizácii a zlepšeniu podnikových procesov.



**Obrázok 3.3-2 LLM poskytuje používateľom možnosť písat kód a získavať výsledky bez potreby programátoriských zručností.**

Podľa štúdie, ktorú uskutočnila spoločnosť Wakefield Research a sponzorovala spoločnosť SAP v roku 2024 [36] a ktorá sa týkala 300 vrcholových manažérov v spoločnostiach s ročnými príjimami najmenej 1 miliardu USD v USA: 52% vrcholových manažérov dôveruje umelej inteligencii pri analýze údajov a poskytovaní odporúčaní pre rozhodovanie. Ďalších 48% využíva AI na identifikáciu predtým nezohľadnených rizík a 47% využíva AI na navrhovanie alternatívnych plánov. Okrem toho 40% využíva AI na vývoj nových produktov, plánovanie rozpočtu a prieskum trhu. Štúdia tiež ukázala pozitívny vplyv AI na osobný život, pričom 39% respondentov uviedlo zlepšenie rovnováhy medzi pracovným a súkromným životom, 38% zlepšenie duševného zdravia a 31% nižšiu úroveň stresu.

Napriek svojej sile však LLM zostávajú nástrojom, ktorý je dôležité používať vedome. Ako každá technológia majú svoje obmedzenia. Jedným z najznámejších problémov sú takzvané "halucinácie" - prípady, keď model s istotou vytvorí hodnovernú, ale fakticky nesprávnu odpoveď. Preto je veľmi dôležité pochopiť, ako model funguje: aké údaje a dátové modely dokáže bezchybne interpretovať, ako interpretuje dotazy a odkiaľ získava informácie. Treba si tiež uvedomiť, že znalosti LLM sú obmedzené dátumom jeho školenia a bez prepojenia na externé údaje nemusí model zohľadňovať aktuálne normy, štandardy, ceny alebo technológie.

Riešením týchto problémov je pravidelná aktualizácia vektorových databáz, pripojenie k relevantným zdrojom a vývoj autonómnych agentov s umelou inteligenciou, ktorí nielen odpovedajú na otázky, ale proaktívne využívajú údaje na školenie, riadia úlohy, identifikujú riziká, navrhujú možnosti optimalizácie a monitorujú výkonnosť projektu.

Prechod na LLM -rozhrania v stavebníctve nie je len technologickou novinkou. Je to zmena paradigmy, ktorá odstraňuje bariéry medzi ľuďmi a údajmi. Je to možnosť pracovať s informáciami rovnako jednoducho, ako sa rozprávame medzi sebou - a pritom získavať presné, overené a akcieschopné výsledky.

Tie spoločnosti, ktoré začnú používať takéto nástroje skôr ako ostatné, získajú významnú konkurenčnú výhodu. Tá zahŕňa zrýchlenie práce, zníženie nákladov a zlepšenie kvality konštrukčných riešení vďaka

rýchlemu prístupu k analýze údajov a schopnosti rýchlo nájsť odpovede na zložité otázky. Je však potrebné zvážiť aj otázky bezpečnosti. Používanie clouдовých služieb LLM -služieb môže byť spojené s rizikami úniku údajov. Preto organizácie čoraz častejšie hľadajú alternatívne riešenia, ktoré im umožnia nasadiť nástroje LLM vo vlastnej infraštruktúre - lokálne, s plnou ochranou a kontrolou nad informáciami.

## Používanie miestnych LLM pre citlivé firemné údaje

Objavenie sa prvých chat-LLM v roku 2022 znamenalo novú etapu vo vývoji umelej inteligencie. Bezprostredne po rozšírení týchto modelov však vznikla oprávnená otázka: Ako bezpečný je prenos údajov a dotazov súvisiacich s podnikom do clodu? Väčšina cloudových jazykových modelov ukladala históriu komunikácie a nahrané dokumenty na svoje servery a pre spoločnosti, ktoré pracujú s citlivými informáciami, to bola vážna prekážka prijatia umelej inteligencie.

Jedným z najudržateľnejších a najlogickejších riešení tohto problému bolo nasadenie Open Source LLM lokálne, v rámci podnikovej IT infraštruktúry. Na rozdiel od cloudových služieb fungujú lokálne modely bez pripojenia na internet, neprenášajú údaje na externé servery a poskytujú podnikom plnú kontrolu nad informáciami

*Najlepší otvorený model [Open Source LLM] je v súčasnosti výkonnostne porovnateľný s uzavretými modelmi [napríklad ChatGPT, Claude], ale s oneskorením približne jedného roka [77].*

- Ben Cottier, vedúci výskumný pracovník neziskovej výskumnej organizácie Epoch AI, 2024

Veľké technologické spoločnosti začali sprístupňovať svoje LLM na miestne použitie. Príkladom prechodu na otvorenú architektúru bol open source rad LLaMA spoločnosti Meta a rýchlo sa rozvíjajúci projekt DeepSeek z Číny. Popri nich aj spoločnosti Mistral a Falcon vydali výkonné modely osloboodené od obmedzení proprietárnych platform. Tieto iniciatívy nielenže urýchliли vývoj globálnej umelej inteligencie, ale poskytli spoločnostiam, ktoré dbajú na ochranu súkromia, skutočné alternatívy nezávislosti, flexibility a dodržiavania bezpečnostných predpisov.

V podnikovom prostredí, najmä v stavebnictve, nie je ochrana údajov len otázkou pohodlia, ale aj dodržiavania právnych predpisov. Práca s dokumentmi z verejných súťaží, odhadmi, výkresmi a dôverou korešpondenciou si vyžaduje prísnu kontrolu. A práve tu poskytuje miestna spoločnosť LLM potrebnú záruku, že údaje zostanú vo vnútri perimetra spoločnosti.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
<b>Data Control</b>	Data is transmitted to third parties	Data remains within the company's network
<b>License</b>	Proprietary, paid	Open-source (Apache 2.0, MIT)
<b>Infrastructure</b>	Requires internet	Operates in an isolated environment
<b>Customization</b>	Limited	Full adaptation to company needs
<b>Cost</b>	Pay-per-token/request	One-time hardware investment + maintenance costs
<b>Scalability</b>	Easily scalable with cloud resources	Scaling requires additional local hardware
<b>Security &amp; Compliance</b>	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
<b>Performance &amp; Latency</b>	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
<b>Integration</b>	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
<b>Updates &amp; Maintenance</b>	Automatically updated by provider	Requires manual updates and model retraining
<b>Energy Consumption</b>	Energy cost is covered by provider	High power consumption for inference and training
<b>Offline Availability</b>	Not available without an internet connection	Works completely offline
<b>Inference Cost</b>	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

Obr. 3.3-3 Lokálne modely poskytujú úplnú kontrolu a bezpečnosť, zatiaľ čo clouдовé riešenia ponúkajú jednoduchú integráciu a automatické aktualizácie.

Hlavné výhody miestneho otvoreného zdroja LLM:

- Úplná kontrola nad údajmi. Všetky informácie zostávajú v rámci spoločnosti, čím sa eliminuje neoprávnený prístup a únik údajov.
- Prevádzka offline. Žiadna závislosť od internetového pripojenia, čo je dôležité najmä pre izolované IT infraštruktúry. Tým sa zabezpečuje nepretržitá prevádzka aj v prípade sankcií alebo blokovania clouдовých služieb.
- Flexibilita aplikácie. Model možno použiť na generovanie textu, analýzu údajov, písanie programového kódu, podporu návrhu a riadenie obchodných procesov.
- Prispôsobenie podnikovým cieľom. LLM možno zaškoliť na interné dokumenty, čo umožňuje zohľadniť špecifickú prácu spoločnosti a jej odvetvové charakteristiky. Miestny LLM možno

prepojiť s platformami CRM, ERP alebo BI, čo vám umožní automatizovať analýzu zákazníckych dopytov, vytváranie prehľadov alebo dokonca prognózovanie trendov.

Nasadenie bezplatného modelu DeepSeek s otvoreným zdrojovým kódom -R1-7B na serveri pre prístup celého tímu používateľov za 1000 USD mesačne môže potenciálne stáť menej ako ročné poplatky za cloudové API, ako napríklad ChatGPT alebo Claude, a umožňuje spoločnostiam prevziať plnú kontrolu nad svojimi údajmi, eliminuje ich prenos na internet a pomáha dodržiavať regulačné požiadavky, ako napríklad GDPR

V iných odvetviach už miestne LLM menia svoj prístup k automatizácii. V oblasti podporných služieb reagujú na časté otázky zákazníkov, čím znižujú pracovné zaťaženie operátorov. V oddeleniach ľudských zdrojov analyzujú životopisy a vyberajú vhodných kandidátov. V elektronickom obchode generujú personalizované ponuky bez toho, aby odhalili údaje používateľa.

Podobný účinok sa očakáva aj v sektore stavebníctva. Integráciou LLM s projektovými údajmi a normami je možné urýchliť prípravu dokumentácie, automatizovať prípravu odhadov a prediktívnych analýz nákladov. Obzvlášť slúbenou oblasťou sa stáva využitie LLM v spojení so štruktúrovanými tabuľkami a dátovými rámcami.

## Úplná kontrola nad AI vo firme a spôsob nasadenia vlastného LLM

Moderné nástroje umožňujú firmám nasadiť veľký jazykový model (LLM) lokálne v priebehu niekoľkých hodín. To poskytuje úplnú kontrolu nad údajmi a infraštruktúrou, eliminuje závislosť od externých clouдовých služieb a minimalizuje riziko úniku informácií. Toto riešenie je obzvlášť dôležité pre organizácie pracujúce s citlivou projektovou dokumentáciou alebo komerčne citlivými údajmi.

V závislosti od úloh a zdrojov sú k dispozícii rôzne scenáre nasadenia, od hotových riešení až po flexibilnejšie a škálovateľné architektúry. Jedným z najjednoduchších nástrojov je Ollama, ktorý umožňuje spustiť jazykové modely doslova na jedno kliknutie bez potreby hlbokých technických znalostí. Rýchly začiatok s aplikáciou Ollama:

1. Stiahnite si distribúciu pre svoj operačný systém (Windows / Linux / macOS) z oficiálnej webovej stránky: [ollama.com](https://ollama.com)
2. Nainštalujte model pomocou príkazového riadku. Napríklad pre model *Mistral*:

```
ollama run mistral
```

3. Po spustení modelu je pripravený na prácu - môžete posieláť textové dotazy cez terminál alebo ho integrovať do iných nástrojov. Spusťte model a vykonajte dotaz:

```
ollama run mistral "Ako vytvoriť výpočet so všetkými zdrojmi pre prácu na inštaláciu 100 mm širokej sadrokartónovej priečky?".
```

Pre tých, ktorí uprednostňujú prácu v známom vizuálnom prostredí, je tu LM Studio, bezplatná aplikácia s rozhraním pripomínajúcim ChatGPT

- Nainštalujte si LM Studio stiahnutím distribučnej sady z oficiálnej webovej stránky - lmstudio.ai
- Prostredníctvom zabudovaného katalógu vyberte model (napr. Falcon alebo GPT-Neo-X) a stiahnite ho.
- Práca s modelom prostredníctvom intuitívneho rozhrania pripomínajúceho ChatGPT, ktoré je však úplne lokalizované.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7-70	16-48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7-13	8-16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7-40	8-32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Obrázok 3.3-4 Porovnanie populárnych lokálnych open source LLM -modelov.

Výber modelu závisí od požiadaviek na rýchlosť, presnosť a dostupné hardvérové možnosti (obrázok 3.3-4). Malé modely, ako napríklad Mistral 7B a Baichuan 7B, sú vhodné na nenáročné úlohy a mobilné zariadenia, zatiaľ čo výkonné modely, ako napríklad DeepSeek -V3, vyžadujú značné výpočtové zdroje, ale ponúkajú vysoký výkon a podporu viacerých jazykov. V nasledujúcich rokoch bude trh s LLM rýchlo rástať - budeme vidieť čoraz viac ľahkých a špecializovaných modelov. Namiesto univerzálnych LLM, ktoré pokrývajú celý ľudský obsah, sa objavia modely vyškolené na základe odborných znalostí z úzkych oblastí. Môžeme napríklad očakávať, že sa objavia modely určené výlučne na spracovanie inžinierskych výpočtov, stavebných odhadov alebo údajov vo formátoch CAD. Takéto špecializované modely budú rýchlejšie, presnejšie a bezpečnejšie na používanie - najmä v profesionálnom prostredí,

kde je dôležitá vysoká spoľahlivosť a hĺbka predmetu.

Po spustení miestneho systému LLM ho možno prispôsobiť konkrétnym úlohám spoločnosti. Na tento účel sa používa technika dolaďovania, pri ktorej sa model ďalej trénuje na interných dokumentoch, technických pokynoch, vzoroch zmlúv alebo projektovej dokumentácii.

## RAG: Inteligentný LLM -asistenti s prístupom k podnikovým údajom

Ďalšou etapou vývoja aplikácie LLM v podnikaní je integrácia modelov so skutočnými podnikovými údajmi v reálnom čase. Tento prístup sa nazýva RAG (Retrieval-Augmented Generation) - vyhľadávanie a rozšírené generovanie. V tejto architektúre sa jazykový model nestáva len dialógovým rozhraním, ale plnohodnotným inteligentným asistentom schopným navigovať v dokumentoch, výkresoch, databázach a poskytovať presné, kontextové odpovede.

Hlavnou výhodou modelu RAG je možnosť využívať interné údaje spoločnosti bez potreby predbežného školenia modelu pri zachovaní vysokej presnosti a flexibility pri spracovaní informácií.

Technológia RAG kombinuje dve hlavné zložky:

- **Vyhľadávanie:** model sa pripája k dátovým úložiskám - dokumenty, tabuľky, PDF - súbory, výkresy - a vyhľadáva relevantné informácie podľa požiadaviek používateľa.
- **Rozšírené generovanie:** na základe získaných údajov model generuje presnú a informovanú odpoveď, pričom zohľadňuje kontext a špecifickosť dopytu.

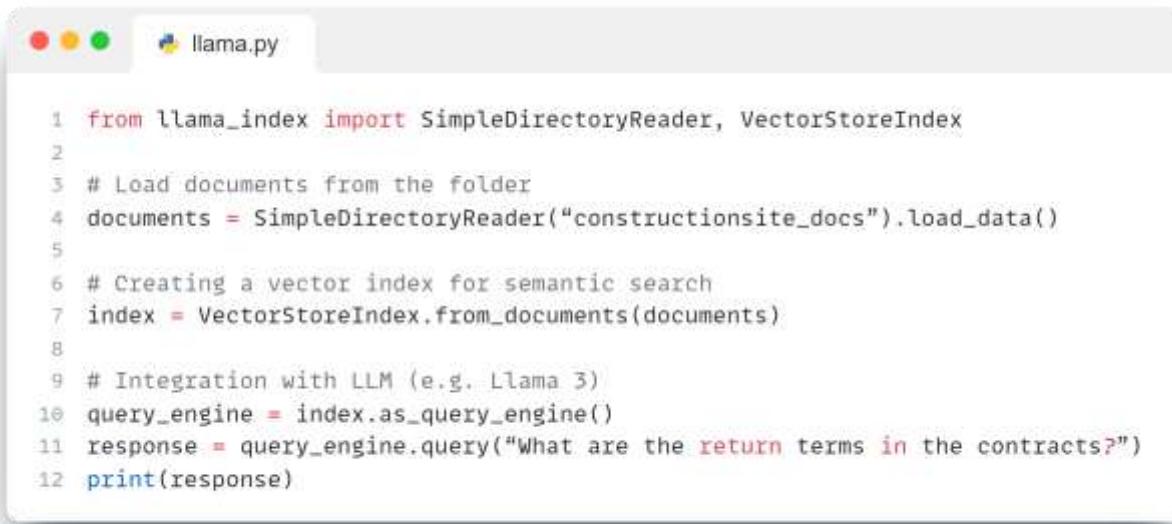
Ak chcete spustiť LLM s podporou RAG, je potrebné vykonať niekoľko krokov:

- **Príprava údajov:** zhromaždenie potrebných dokumentov, výkresov, špecifikácií, tabuľiek. Môžu byť v rôznych formátoch a štruktúrach, od PDF až po Excel.
- **Indexovanie a vektorizácia:** pomocou nástrojov, ako je LlamaIndex alebo LangChain, sa údaje konvertujú na vektorové reprezentácie, ktoré umožňujú nájsť sémantické prepojenia medzi textovými fragmentmi (viac o vektorových databázach a prevode veľkých polí na vektorovú reprezentáciu vrátane projektov CAD v časti 8).
- **Dotazovanie asistenta:** po nahráti údajov môžete modelu klásiť otázky a on bude hľadať odpovede v rámci podniku, a nie vo všeobecných znalostiah získaných z internetu.

Predpokladajme, že spoločnosť má priečinok constructionsite\_docs, v ktorom sú uložené zmluvy, pokyny, odhady a tabuľky. Pomocou skriptu v jazyku Python (obr. 3.3-5) môžeme skenovať tento priečinok a vytvoriť vektorové indexovanie: každý dokument sa prevedie na súbor vektorov odrážajúcich sémantický obsah textu. Tým sa dokumenty zmenia na akúsi "mapu významov", na ktorej sa model môže efektívne pohybovať a nachádzať súvislosti medzi termínmi a frázami.

Model si napríklad "pamäta", že slová "vrátenie" a "reklamácia" sa často nachádzajú v časti zmluvy

týkajúcej sa prepravy materiálov na stavenisko. Potom, ak je položená otázka - napríklad "Aká je naša lehota na vrátenie?", môže sa stať, že sa objaví otázka, ktorá sa týka vrátenia tovaru. (Obr. 3.3-5 - riadok 11 kódu) - LLM analyzuje interné dokumenty a nájde presné informácie, pričom sa správa ako inteligentný asistent schopný prečítať a pochopiť obsah všetkých firemných súborov.



```

1 from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3 # Load documents from the folder
4 documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6 # Creating a vector index for semantic search
7 index = VectorStoreIndex.from_documents(documents)
8
9 # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

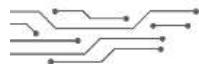
```

Obr. 3.3-5 LM číta priečinok so súbormi - podobne ako keď ho človek otvorí a hľadá požadovaný dokument

Kód je možné spustiť na akomkoľvek počítači s nainštalovaným jazykom Python. O používaní jazyka Python a IDE na spúšťanie kódu si povieme viac v nasledujúcej kapitole.

Lokálne nasadenie LLM nie je len trendom, ale strategickým riešením pre spoločnosti, ktoré si cenia bezpečnosť a flexibilitu. Nasadenie LLM, či už na lokálnych firemných počítačoch alebo pomocou online riešení, je však len prvým krokom. Na to, aby mohli spoločnosti uplatniť možnosti LLM na reálne úlohy, musia používať nástroje, ktoré im umožnia nielen prijímať odpovede z chatu, ale aj ukladať vytvorenú logiku vo forme kódu, ktorý možno spustiť mimo kontextu používania LLM. To je dôležité pre škálovanie riešení - správne organizované procesy umožňujú aplikovať vývoj AI na niekoľko projektov naraz alebo dokonca v rámci celej spoločnosti.

V tejto súvislosti zohráva dôležitú úlohu výber vhodného vývojového prostredia (IDE). Moderné programovacie nástroje umožňujú nielen vyvíjať riešenia založené na LLM, ale aj integrovať ich do existujúcich podnikových procesov a premeniť ich na automatizované ETL-Pipeline



## KAPITOLA 3.4.

### IDE S PODPOROU LLM A BUDÚCE PROGRAMOVÉ ZMENY

#### Výber IDE: od experimentov LLM po obchodné riešenia

Pri ponáraní sa do sveta automatizácie, analýzy údajov a umelej inteligencie - najmä pri práci s veľkými jazykovými modelmi (LLM) - je veľmi dôležité vybrať si správne integrované vývojové prostredie (IDE). Toto IDE bude vaším hlavným pracovným nástrojom: miestom, kde sa bude spúštať kód vygenerovaný LLM, a to na lokálnom počítači aj v rámci podnikovej siete. Výber IDE určuje nielen pohodlie vašej práce, ale aj to, ako rýchlo budete môcť prejsť od experimentálnych dotazov LLM k plnohodnotným riešeniam začleneným do reálnych podnikových procesov.

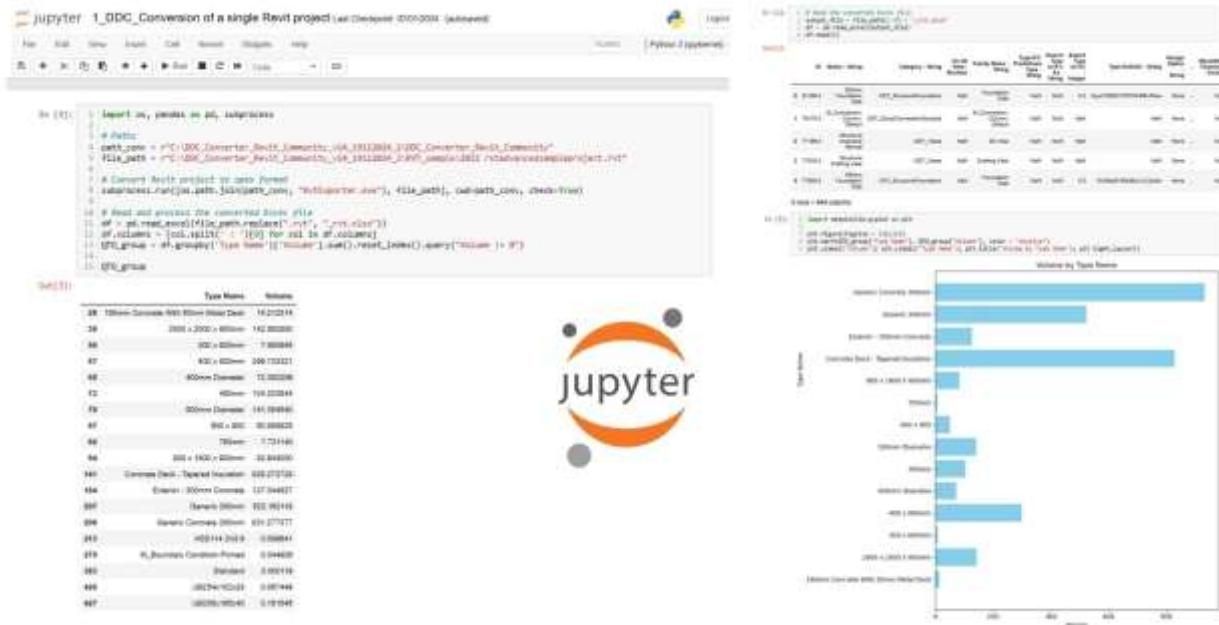
**IDE** (integrované vývojové prostredie) je univerzálny stavebný prvk v počítači na automatizáciu procesov a spracovanie údajov. Namiesto toho, aby ste mali samostatne pílu, kladivo, vrtačku a ďalšie nástroje, máte jedno zariadenie, ktoré dokáže všetko - rezat', upevňovať, vrtať a dokonca aj kontrolovať kvalitu materiálov. IDE pre programátorov je jednotný priestor, kde môžete písat kód (analogicky s konštrukciou - vytvárať výkresy), testovať jeho prácu (montáž stavebného modelu), hľadať chyby (ako pri kontrole pevnosti konštrukcií pri stavbe) a spustiť hotový projekt (uvedenie domu do prevádzky).

Prehľad populárnych IDE:

- **PyCharm®** (JetBrains) je výkonné profesionálne IDE pre Python. Vďaka veľkému množstvu zabudovaných funkcií je vhodné na seriózne projekty. Základná podpora interaktívnych súborov Jupyter (IPYNB) je však k dispozícii len v platenej verzii a pre začiatočníkov môže byť rozhranie neprehľadné.

Súbor s príponou IPYNB (Interactive Python Notebook) je formát pre interaktívne zápisníky Jupyter® (obrázok 3.4-1), v ktorom sú kód, vizualizácie a vysvetlenia spojené do jedného dokumentu. Tento formát je ideálny na vytváranie správ, analytických a trénin-gových scenárov.

- **VS Code®** (Microsoft) je rýchly, flexibilný a prispôsobiteľný nástroj s bezplatnou podporou IPYNB a mnohými doplnkami. Je vhodný pre začiatočníkov aj profesionálov. Umožňuje integráciu zásuvných modulov GitHub Copilot a jazykových modelov, vďaka čomu je skvelou voľbou pre projekty AI a dátovej vedy.
- **Zápisník Jupyter** - klasická a oblúbená voľba na experimentovanie a učenie. Umožňuje písat kód, pridávať vysvetlenia a vizualizovať výsledky v jednom rozhraní (obr. 3.4-1). Ideálny na rýchle testovanie hypotéz, prácu s LLM a vytváranie reprodukovateľných krokov analýzy di-vokých údajov. Na správu závislostí a knižníc odporúčame používať Anaconda Navigator, vizuálne rozhranie na správu prostredia Python.



Obr. 3.4-1 Jupyter Notebook jeden z najpohodlniejsích a najobľúbenejších nástrojov na vytváranie procesov Pipeline.

■ **Google Collab™** (a platforma Kaggle (obrázok 9.2-5)) je clouдовá alternatíva k Jupyteru, ktorá poskytuje bezplatný prístup ku GPU/TPU. Je to skvelé riešenie na začiatok - žiadna lokálna inštalácia softvéru a možnosť pracovať priamo z prehliadača. Podporuje integráciu s Diskom Google a nedávno aj s Gemini (LLM spoločnosti Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
<b>Complexity</b>	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

Obrázok 3.4-2 Porovnanie IDE: Jupyter Notebook jeden z najpohodlniejsích a najjednoduchších nástrojov na vytváranie procesov Pipeline.

Výber IDE závisí od vašich úloh. Ak chcete rýchlo začať pracovať s AI, vyskúšajte Jupyter Notebook alebo Google Collab. Na seriózne projekty je lepšie použiť PyCharm alebo VS Code. Najdôležitejšie je začať. Moderné nástroje vám umožnia rýchlo premeniť vaše experimenty na funkčné riešenia.

Všetky vyššie opísané IDE umožňujú vytvárať potrubia na spracovanie údajov - t. j. reťazce modulov blokov kódu (ktoré by mohol generovať LLM), z ktorých každý je zodpovedný napríklad za inú fázu:

- analytické scenáre,
- reťazce získavania informácií z dokumentov,
- automatické reakcie na základe RAG,
- generovanie správ a vizualizácií.

Vďaka modulárnej štruktúre možno každý krok predstaviť ako samostatný blok: načítanie údajov → filtrovanie → analýza → vizualizácia → export výsledkov. Tieto bloky možno opakovane používať, - prispôsobovať a zostavovať do nových reťazcov ako konštruktér, len pre údaje.

Inžinierom, manažérom a analytikom sa tak otvára možnosť dokumentovať logiku rozhodovania vo forme kódu, ktorý možno generovať pomocou LLM. Tento prístup pomáha urýchliť rutinné úlohy, automatizovať typické operácie a vytvoriť opakovateľné procesy, v ktorých je každý krok jasne zdokumentovaný a transparentný pre všetkých členov tímu.

Automatizované nástroje ETL Pipelines (obr. 7.2-3), Apache Airflow (obr. 7.4-4), Apache NiFi (obr. 7.4-5) a n8n (obr. 7.4-6na) vytváranie blokov logiky pre automatizáciu procesov budú podrobnejšie rozobrané v 7. a 8. časti knihy.

## **IDE s podporou LLM a budúce zmeny v programovaní**

Integrácia umelej inteligencie do vývojových procesov mení prostredie programovania. Moderné prostredia už nie sú len textové editory so zvýrazňovaním syntaxe - menia sa na inteligentných asistentov schopných pochopiť logiku projektu, doplniť kód a dokonca vysvetliť, ako funguje konkrétny fragment kódu. Na trhu sa objavujú produkty, ktoré využívajú umelú inteligenciu na posúvanie hraníc bežného vývoja:

- **GitHub Copilot** (integruje sa s VS Code, PyCharm): AI -asistent, ktorý generuje kód na základe komentárov alebo čiastočných opisov a mení textové nápovedy na hotové riešenia.
- **Cursor** (fork VS Code s AI -kernel): umožňuje nielen dokončiť písanie kódu, ale aj klášť otázky projektu, hľadať závislosti a učiť sa zo základne kódu.
- **JetBrains AI Assistant**: zásuvný modul pre JetBrains IDE (vrátane PyCharm) s funkciou vysvetľovania zložitého kódu, optimalizácie a tvorby testov.
- **Amazon CodeWhisperer**: obdoba Copilota so zameraním na bezpečnosť a podporu služieb Amazon AWS.

Programovanie sa v nadchádzajúcich rokoch dramaticky zmení. Hlavné zameranie sa presunie z rutinného písania kódu na návrh modelu a dátovej architektúry - vývojári sa budú viac podieľať na návrhu systému, zatiaľ čo AI prevezme šablónové úlohy: generovanie kódu, testy, dokumentáciu a základné funkcie. Budúcnosť programovania je spolupráca medzi ľuďmi a umelou inteligenciou, kde stroje preberú technickú rutinu a ľudia sa zamerajú na kreativitu.

Programovanie v prirodzenom jazyku sa stane každodenným javom. Personalizácia IDE dosiahne novú

úroveň - vývojové prostredia sa naučia prispôsobiť štýlu práce používateľa a jeho spoločnosti tým, že budú predvídať vzory, ponúkať kontextové riešenia a učiť sa z predchádzajúcich projektov

Úloha vývojára sa tým nezruší, ale radikálne sa zmení: od písania kódu k riadeniu znalostí, kvality a procesov. Tento vývoj ovplyvní aj business intelligence, kde sa tvorba reportov, vizualizácií a aplikácií na podporu rozhodovania bude čoraz viac realizovať prostredníctvom generovania kódu a logiky s pomocou AI a LLM, chatu a agentových rozhraní.

Po zriadení chatov LLM a výbere vhodného vývojového prostredia je ďalším dôležitým krokom organizácia údajov. Tento proces zahŕňa extrakciu informácií z rôznorodých zdrojov, ich čistenie, transformáciu do štruktúrovanej podoby a integráciu do podnikových systémov.

V modernom dátovo orientovanom prístupe k správe údajov je hlavným cieľom priniesť údaje do jednej univerzálnej podoby, ktorá je kompatibilná s veľkým množstvom nástrojov a aplikácií. Na spracovanie štruktúrovaných procesov a štruktúrovaných údajov sú potrebné špecializované knižnice. Jednou z najvýkonnejších, najflexibilnejších a najobľúbenejších je knižnica Pandas pre Python. Umožňuje pohodlne spracovávať tabuľkové údaje: filtrovať, zoskupovať, čistiť, pripájať, agregovať a vytvárať zostavy.

## Python Pandas: nenahraditeľný nástroj na prácu s údajmi

Pandas zaujíma vo svete analýzy a automatizácie údajov osobitné miesto. Je to jedna z najpopulárnejších a najpoužívanejších knižníc programovacieho jazyka Python [86], určená na prácu so štruktúrovanými údajmi.

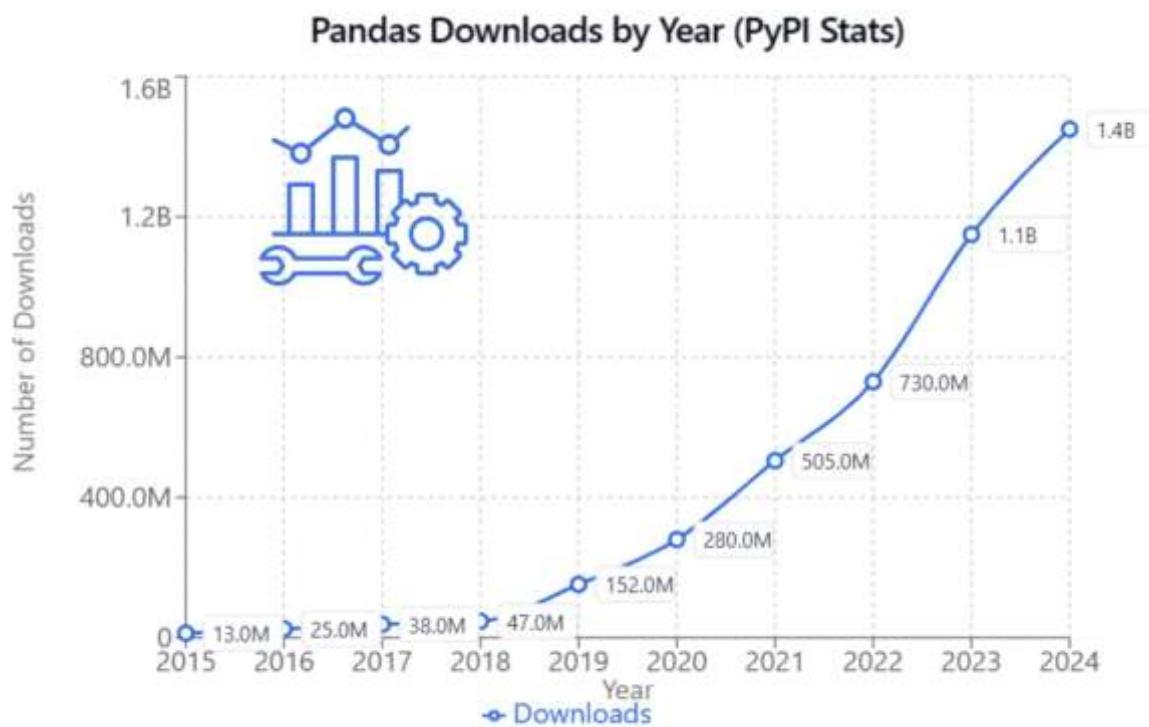
Knižnica je ako súbor hotových nástrojov: funkcií, modulov, tried. Tak ako na stavbe nemusíte každým vymýšlať kladivo alebo vodováhu, tak aj v programovaní vám knižnica umožňujú rýchlo riešiť problémy bez toho, aby ste museli znova vymýšlať základné funkcie a riešenia.

Pandas je open source knižnica jazyka Python , ktorá poskytuje vysoko výkonné a intuitívne dátové štruktúry, najmä DataFrame, univerzálny formát na prácu s tabuľkami. Pandas je švajčiarsky nôž pre analytikov, inžinierov a vývojárov pracujúcich s údajmi.

Python je vysokoúrovňový programovací jazyk s jednoduchou syntaxou, ktorý sa aktívne používa v analytike, automatizácii, strojovom učení a pri vývoji webových stránok. Jeho popularita je spôsobená čitateľnosťou kódu, multiplatformovosťou a bohatým ekosystémom knižníc. K dnešnému dňu bolo pre Python vytvorených viac ako 137 000 balíkov s otvoreným zdrojovým kódom [87] a tento počet takmer denne narastá. Každá takáto knižnica je akýmsi úložiskom hotových funkcií: od jednoduchých matematických operácií až po komplexné nástroje na spracovanie obrazu, analýzu veľkých dát, neurónové siete a integráciu s externými službami.

Inými slovami, predstavte si, že máte bezplatný a otvorený prístup k stovkám tisícok hmotívych softvérových riešení - knižníc a nástrojov, ktoré môžete priamo začleniť do svojich podnikových procesov. Je to ako obrovský katalóg aplikácií na automatizáciu, analýzu, vizualizáciu, integráciu a ďalšie - a všetky sú k dispozícii okamžite po nainštalovaní jazyka Python.

Pandas je jedným z najpopulárnejších balíkov v ekosystéme Python. V roku 2022 dosiahol priemerný počet stiahnutí knižnice Pandas 4 milióny za deň (obrázok 3.4-3), zatiaľ čo začiatkom roka 2025 sa toto číslo zvýšilo na 12 miliónov stiahnutí za deň, čo odráža jej rastúcu popularitu a široké využitie v dátovej analytike a LLM chatoch [86]

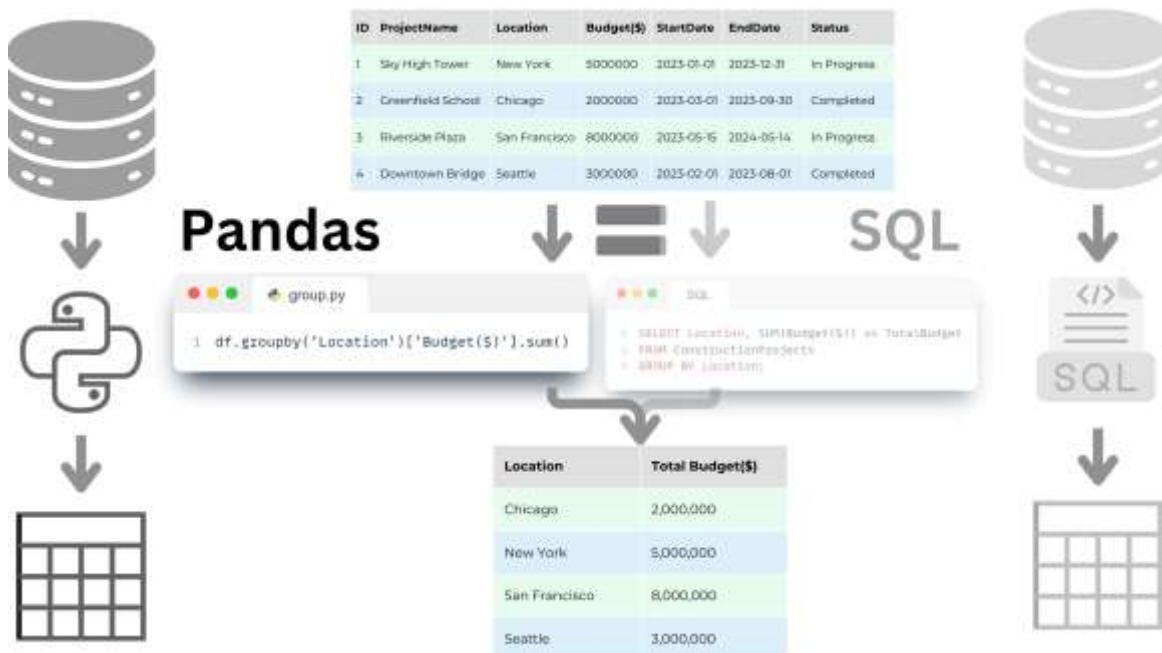


Obrázok 3.4-3 Pandas je jednou z najviac stiahovaných knižníc. V roku 2024 jej ročný počet stiahnutí presiahol 1,4 miliardy.

Dotazovací jazyk v knižnici Pandas je svojou funkčnosťou podobný dotazovaciemu jazyku SQL, o ktorom sme hovorili v kapitole "Relačné databázy a dotazovací jazyk SQL".

Vo svete analytiky a správy štruktúrovaných údajov vyniká Pandas svojou jednoduchosťou, rýchlosťou a výkonom a poskytuje používateľom širokú škálu nástrojov na efektívnu analýzu a spracovanie informácií.

Oba nástroje - SQL a Pandas - poskytujú výkonné možnosti manipulácie s údajmi, najmä v porovnaní s tradičným programom Excel. Podporujú operácie, ako je výber, filtrovanie (obrázok 3.4-4), pričom jedený rozdiel je v tom, že SQL je optimalizovaný na prácu s relačnými databázami, zatiaľ čo Pandas spracováva údaje v pamäti RAM, čo mu umožňuje bežať na akomkoľvek počítači bez potreby vytvárania databáz a nasadzovania samostatnej infraštruktúry.

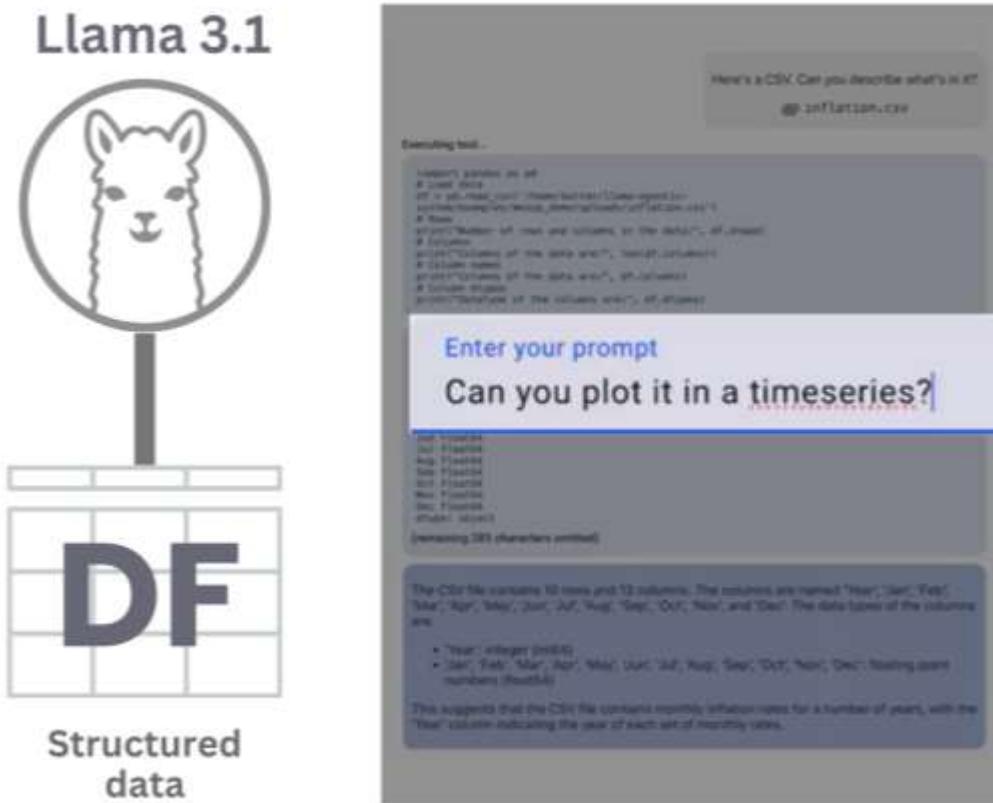


Obr. 3.4-4 Pandas, na rozdiel od SQL, má flexibilitu na prácu s rôznymi dátovými formátmi, neobmedzuje sa len na databázy.

Pandas sa často uprednostňuje na vedecký výskum, automatizáciu procesov, vytváranie pipeline (vrátane ETL) a manipuláciu s údajmi v jazyku Python, zatiaľ čo SQL je štandardom správy databáz a často sa používa v podnikových prostrediach na spracovanie veľkého množstva údajov.

Knižnica Pandas programovacieho jazyka Python umožňuje vykonávať nielen základné operácie, ako je čítanie a zápis tabuľiek, ale aj zložitejšie úlohy vrátane zlučovania údajov, zoskupovania údajov a vykonávania zložitých analytických výpočtov.

Dnes sa knižnica Pandas používa nielen v akademickom výskume a podnikovej analytike, ale aj v spojení s LLM -modelmi. Napríklad divízia Meta® (Facebook™) pri vydaní nového open source modelu LLaMa 3.1 v roku 2024 venovala osobitnú pozornosť práci so štruktúrovanými údajmi, pričom jedným z kľúčových a prvých prípadov v jeho vydaní bolo práve spracovanie štruktúrovaných dátových rámcov (obr. 3.4-5) vo formáte CSV a integrácia s knižnicou Pandas priamo v chate.



Obr. 3.4-5 Jedným z prvých a hlavných prípadov tímu Meta, ktorý bol prezentovaný v LlaMa 3.1 v roku 2024, bolo vytváranie aplikácií pomocou Pandas.

Pandas je základným nástrojom pre milióny dátových vedcov, ktorí spracovávajú a pripravujú údaje pre generatívnu AI. Zrýchlenie Pandas s nulovými zmenami kódu bude obrovským krokom vpred. Dátoví vedci budú môcť spracovať údaje v priebehu niekoľkých minút namiesto hodín a získajú rádovo viac údajov na trénovanie generatívnych modelov AI [88].

- Jensen Huang, zakladateľ a generálny riaditeľ spoločnosti NVIDIA

Pomocou programu Pandas, je možné spravovať a analyzovať súbory údajov, ktoré daleko presahujú možnosti programu Excel. Zatiaľ čo Excel je zvyčajne schopný spracovať maximálne 1 milión riadkov údajov, Pandas dokáže ľahko spracovať súbory údajov (obrázok 9.1-2, obrázok 9.1-10) obsahujúce desiatky miliónov riadkov [89]. Táto schopnosť umožňuje používateľom vykonávať sofistikovanú analýzu a vizualizáciu údajov na veľkých súboroch údajov, čo poskytuje hlboký prehľad a uľahčuje rozhodovanie založené na údajoch. Okrem toho má Pandas silnú podporu komunity [90]: stovky miliónov vývojárov a analytikov na celom svete (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) ho používajú online alebo offline každý deň, čo poskytuje veľké množstvo hotových riešení pre akýkoľvek obchodný problém.

Základom väčšiny analytických procesov v jazyku Python je štruktúrovaná forma údajov nazývaná DataFrame, ktorú poskytuje knižnica Pandas. Je to výkonný a flexibilný nástroj na organizáciu, analýzu a vizualizáciu tabuľkových údajov.

## DataFrame: univerzálny formát tabuľkových údajov

DataFrame je centrálna štruktúra v knižnici Pandas, ktorá predstavuje dvojrozmernú tabuľku (obrázok 3.4-6), kde riadky zodpovedajú jednotlivým objektom alebo záznamom a stĺpce zodpovedajú ich vlastnostiam, parametrom alebo kategóriám. Táto štruktúra vizuálne pripomína tabuľky Excelu, ale je oveľa lepšia z hľadiska flexibility, škálovateľnosti a funkčnosti.

**DataFrame** je spôsob reprezentácie a spracovania tabuľkových údajov uložených v pamäti RAM počítača.

DataFrame je spôsob reprezentácie a spracovania tabuľkových údajov uložených v pamäti RAM počítača. V tabuľke môžu riadky vyjadrovať napríklad prvky stavebného projektu a stĺpce ich vlastnosti: kategórie, rozmery, súradnice, náklady, termíny atď. Navyše takáto tabuľka môže obsahovať informácie o jednom projekte (obr. 4.1-13) aj údaje o miliónoch objektov z tisícov rôznych projektov (obr. 9.1-10). Vďaka vektorizovaným operáciám Pandas je možné takéto objemy informácií ľahko filtrovať, zoskupovať a agregovať vysokou rýchlosťou.

The diagram illustrates a DataFrame as a structured data table. It features a building crane icon at the top left. To the right, the text "STRUCTURED DATA" is displayed next to a grid icon. Below this, a table is shown with the following columns: ID, Name, Category, Family Name, Height, BoundingBoxMin\_X, BoundingBoxMin\_Y, BoundingBoxMin\_Z, and Level. The table contains several rows of data. Annotations point to specific parts of the table:

- "Index label" points to the first row (index 0).
- "Index axis = 0" points to the first column.
- "Missing value" points to an empty cell in the table.
- "Data" points to the main body of the table.
- "Column names" points to the header row.

Obrázok 3.4-6 Stavebný projekt ako dátový rámec je dvojrozmerná tabuľka s prvkami v riadkoch a atribútmi v stĺpcoch.

Spoločnosť Nvidia odhaduje, že už dnes sa až 30% všetkých výpočtových zdrojov využíva na spracovanie štruktúrovaných údajov - dátových rámcov - a tento podiel nadálej rastie.

*Spracovanie údajov je to, čo sa v každej spoločnosti vykonáva pravdepodobne jedna tretina svetovej výpočtovej techniky. Spracovanie údajov a dátá väčšiny spoločností sú v DataFrame, vo formáte tabuľky*

- Jensen Huang, generálny riaditeľ spoločnosti Nvidia [91]

Uvedme si niektoré kľúčové vlastnosti DataFrame v Pandas:

- Stĺpce: v DataFrame sú údaje usporiadane do stĺpcov, z ktorých každý má jedinečný názov. Atribútové stĺpce môžu obsahovať údaje rôznych typov, podobne ako stĺpce v databázach alebo stĺpce v tabuľkách.
- **Pandas Séria** je jednorozmerná dátová štruktúra v programe Pandas, podobná zoznamu alebo stĺpcu v tabuľke, kde každá hodnota zodpovedá inému indexu.

Séria Pandas má viac ako 400 atribútov a metód, vďaka čomu je práca s údajmi neuveriteľne flexibilná. Na stĺpec môžete priamo aplikovať jednu zo štyroch stoviek dostupných funkcií, vykonávať matematické operácie, filtrovať údaje, nahrádať hodnoty, pracovať s dátumami, reťazcami a oveľa viac. Okrem toho Series podporuje vektorové operácie, čo v porovnaní s cyklickými výpočtami výrazne urýchluje spracovanie veľkých súborov údajov. Môžete napríklad jednoducho vynásobiť všetky hodnoty číslom, nahradíť chýbajúce údaje alebo použiť zložité transformácie bez toho, aby ste museli písat zložité cykly.

- **Riadky**: v DataFrame môžu byť indexované jedinečnými hodnotami. Tento index umožňuje rýchlo meniť a upravovať údaje v konkrétnych riadkoch.
- **Index**: Pri vytváraní rámca údajov Pandas sa každému riadku štandardne priradí index od 0 do N-1 (kde N je počet všetkých riadkov v rámci údajov). Index však môžete zmeniť tak, aby obsahoval špeciálne označenia, ako sú dátumy alebo jedinečné charakteristiky.
- **Indexovanie** riadkov v DataFrame znamená, že každému riadku je priradený jedinečný názov alebo označenie, ktoré sa nazýva index DataFrame.
- **Dátové typy**: DataFrame podporuje rôzne dátové typy vrátane: `int`, `float`, `bool`, `datetime64` a `object` pre textové údaje. Každý stĺpec DataFrame má svoj vlastný dátový typ, ktorý určuje, aké operácie možno vykonávať s jeho obsahom.
- **Operácie s údajmi**: DataFrame podporuje širokú škálu operácií na spracovanie údajov vrátane agregácie (`groupby`), zlúčovania (`merge` a `join`), konkatenácie (`concat`), rozdeľenia-aplikácie-kombinácie a mnohých ďalších techník transformácie údajov.
- **Manipulácia s veľkosťou**: DataFrame umožňuje pridávať a odstraňovať stĺpce a riadky, čím sa stáva dynamickou štruktúrou, ktorú možno upravovať podľa potrieb analýzy údajov.
- **Vizualizácia údajov**: Pomocou zabudovaných vizualizačných techník alebo prepojením s populárnymi knižnicami na vizualizáciu údajov, ako sú Matplotlib alebo Seaborn, možno DataFrame ľahko konvertovať na grafy a diagramy na grafickú prezentáciu údajov.
- **Vstup a výstup údajov**: Pandas poskytuje funkcie na čítanie import a export údajov do

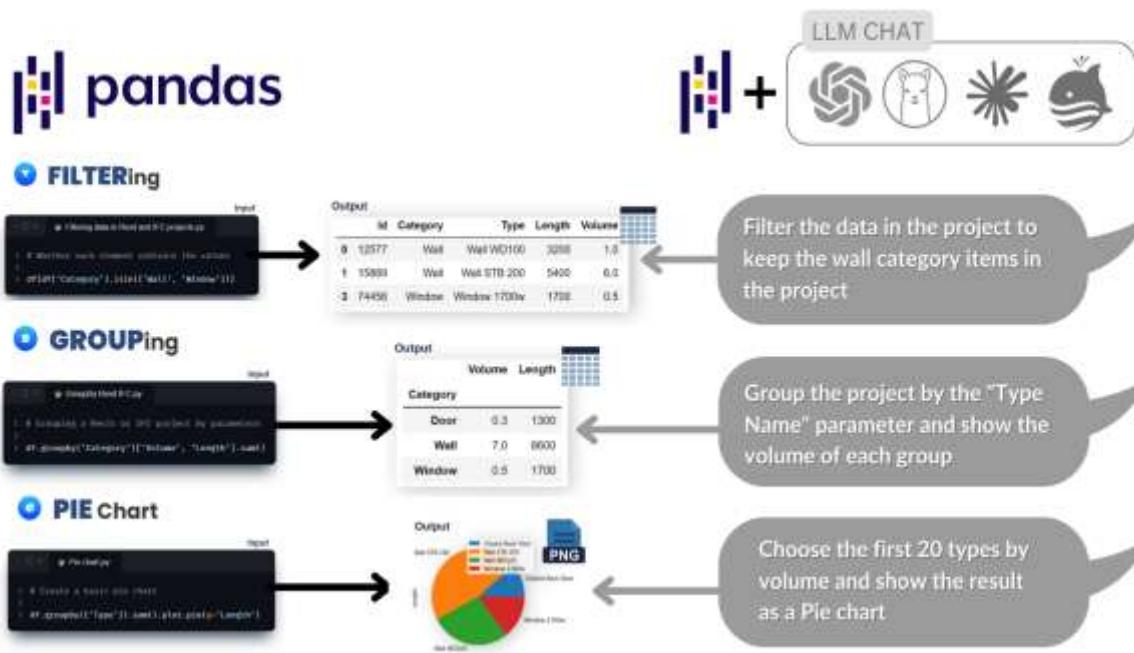
rôznych formátov súborov, ako sú CSV, Excel, JSON, HTML a SQL, čím sa DataFrame môže stať centrálnym centrom na zber a distribúciu údajov.

Na rozdiel od CSV a XLSX, Pandas DataFrame poskytuje väčšiu flexibilitu a výkon pri práci s údajmi: dokáže spracovať veľké množstvo informácií v pamäti RAM, podporuje rozšírené typy údajov (vrátane dátumov, logických hodnôt a časových radov) a poskytuje rozsiahle možnosti filtrovania, agregácie, spájania a vizualizácie údajov. Zatiaľ čo CSV neukladá informácie o typoch a štruktúre údajov a XLSX je často preťažený formátovaním a má nízku škálovateľnosť, DataFrame zostáva optimálnou voľbou na rýchlu analýzu, automatizáciu procesov a integráciu s AI -modelmi (obrázok 3.4-7). Nasledujúce kapitoly sa budú podrobne zaoberať každým z týchto aspektov údajov, v 8. časti knihy sa tiež podrobne rozoberú podobné formáty, ako sú Parquet, Apache Orc, JSON, Feather, HDF5 a dátové sklady (obr. 8.1-2).

		XLSX	CSV	Pandas DataFrame
	Storage	Tabular	Tabular	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
	Compression	Built-in	None	None (in-memory)
	Performance	Low	Medium	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Low
	Data Type Support	Limited	Very limited	Extended
	Scalability	Low	Low	Medium (memory limited)

Obr. 3.4-7 DataFrame je optimálnou voľbou pre manipuláciu s dátami s vysokým výkonom a pokročilou podporou dátových typov.

Knižnica Pandas a formát DataFrame sa vďaka svojej flexibilite, výkonu a jednoduchému používaniu stali de facto štandardom v analýze údajov v jazyku Python. Sú ideálnymi na vytváranie jednoduchých zostáv aj na budovanie komplexných analytických potrubí, najmä v spojení s modelmi LLM.



Obr. 3.4-8 LLM zjednodušuje interakciu s Pandas: namiesto kódu stačí textový dotaz.

Dnes sa Pandas aktívne používa v chatovacích miestnostiach založených na LLM, ako sú ChatGPT, LLaMa, DeepSeek, QWEN a iné. V mnohých prípadoch, keď model dostane dotaz týkajúci sa spracovania tabuľiek, overovania údajov alebo analýzy, generuje kód práve pomocou knižnice Pandas. Vďaka tomu je DataFrame prirodzeným "jazykom" na reprezentáciu údajov v dialógoch s umelou inteligenciou (obrázok 3.4-8).

Moderné dátové technológie, ako napríklad Pandas, uľahčujú analýzu, automatizáciu a integráciu údajov do obchodných procesov. Prinášajú rýchle výsledky, znižujú pracovné zaťaženie špecialistov a zabezpečujú opakovateľné operácie.

## Ďalšie kroky: budovanie udržateľného dátového rámca

V tejto časti sme sa venovali hlavným typom údajov používaných v stavebnictve, dozvedeli sme sa o rôznych formátoch na ich ukladanie a analyzovali sme úlohu moderných nástrojov vrátane LLM a IDE pri spracovaní informácií. Dozvedeli sme sa, že efektívna správa údajov je základom pre informované rozhodovanie a automatizáciu podnikových procesov. Organizácie, ktoré dokážu štruktúrovať a organizovať svoje údaje, získavajú významnú konkurenčnú výhodu vo fázach spracovania a transformácie údajov.

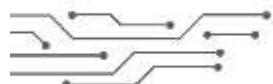
Na zhrnutie tejto časti stojí za to zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované prístupy vo vašich každodenných úlohách:

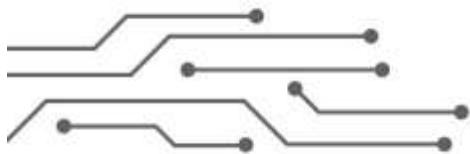
- Vykonajte dátový audit svojich procesov
- Urobte si súpis všetkých typov údajov, ktoré používate vo svojich projektoch

- Určite, ktoré typy údajov a modely sú pre vaše obchodné procesy najdôležitejšie.
- Identifikovať problémové oblasti, v ktorých sú informácie často neštruktúrované, zle štruktúrované alebo nedostupné.
- Začnite vytvárať stratégii správy údajov**
  - Vyzdvihnuť politické otázky a normy pre prácu s rôznymi typmi údajov.
  - Analyzujte, ktoré z vašich pracovných postupov možno zlepšiť konverziou neštruktúrovaných údajov na štruktúrované.
  - Vytvorenie politiky ukladania údajov a prístupu k nim, ktorá zohľadňuje bezpečnosť a dôvernosť
- Inštalácia a ovládanie základných nástrojov na prácu s údajmi**
  - Vyberte si vhodné IDE, ktoré zodpovedá vašim úlohám (napr. nainštalujte VS Code alebo Jupyter Notebook)
  - Skúste si zriadiť miestnu službu LLM, ktorá bude dôverne spracovávať vaše osobné údaje.
  - Začnite experimentovať s knižnicou Pandas na spracovanie tabuľkových údajov XLSX
  - Popíšte LLM typické úlohy, ktoré riešite v tabuľkových nástrojoch alebo databázach, a požiadajte LLM o automatizáciu práce pomocou programu Pandas

Uplatňovanie týchto krokov vám umožní postupne zmeniť prístup k práci s údajmi a prejsť od rôznorodých, neštruktúrovaných súborov údajov k jednotnému ekosystému, v ktorom sa údaje stanú dostupným a zrozumiteľným aktívom. Začnite v malom - vytvorte svoj prvý DataFrame v programe Pandas, spusťte lokálny LLM, zautomatizujte svoju prvú rutinnú úlohu pomocou jazyka Python (napr. tabuľky Excel).

Štvrtá časť knihy sa zameriava na kvalitu údajov, organizáciu, štruktúrovanie a modelovanie údajov. Zameriame sa na metodiky, ktoré transformujú rôznorodé zdroje údajov - od PDF a textov až po obrázky a modely CAD - na štruktúrované súbory údajov vhodné na analýzu a automatizáciu. Budeme tiež skúmať, ako sa formalizujú požiadavky na údaje, ako sa vytvárajú konceptuálne a logické modely v konštrukčných projektoch a ako môžu v tomto procese pomôcť moderné jazykové modely (LLM).





## IV ČASŤ

### **KVALITA ÚDAJOV: ORGANIZÁCIA, ŠTRUKTÚROVANIE, MODELOVANIE**

Štvrtá časť sa zameriava na metodiky a technológie, ktoré zabezpečujú transformáciu rôznorodých informácií na kvalitné štruktúrované súbory údajov. Podrobne sa rozoberajú procesy tvorby a dokumentovania požiadaviek na údaje ako základu efektívnej informačnej architektúry v stavebných projektoch. Uvádzajú sa praktické metódy získavania štruktúrovaných informácií z rôznych zdrojov (PDF -dokumenty, obrázky, textové súbory, CAD -modely) s príkladmi realizácie. Analyzuje sa použitie regulárnych výrazov (RegEx) a ďalších nástrojov na automatickú validáciu a verifikáciu údajov. Postupne sa opisuje proces modelovania údajov na koncepčnej, logickej a fyzickej úrovni, pričom sa zohľadňujú špecifika stavebníctva. Uvádzajú sa konkrétne príklady použitia jazykových modelov (LLM) na automatizáciu procesov štruktúrovania a overovania informácií. Navrhujú sa efektívne prístupy k vizualizácii výsledkov analýz, čím sa zvyšuje dostupnosť analytických informácií pre všetky úrovne riadenia stavebných projektov

## KAPITOLA 4.1.

### PREVOD ÚDAJOV DO ŠTRUKTÚROVANEJ PODOBY

V ére ekonomiky založenej na údajoch sa údaje stávajú základom rozhodovania, a nie prekážkou. Namiesto neustáleho prispôsobovania informácií každému novému systému a jeho formátom sa spoločnosti čoraz viac snažia vytvoriť jednotný štruktúrovaný dátový model, ktorý slúži ako univerzálny zdroj pravdy pre všetky procesy. Moderné informačné systémy sa navrhujú nie na základe formátov a rozhraní, ale na základe významu údajov - pretože štruktúra sa môže meniť, ale význam informácií zostáva rovnaký oveľa dlhšie.

Kľúčom k efektívnej práci s údajmi nie je ich nekonečná konverzia a transformácia, ale ich správna organizácia od začiatku: vytvorenie univerzálnej štruktúry, ktorá dokáže zabezpečiť transparentnosť, automatizáciu a integráciu vo všetkých fázach životného cyklu projektu.

Tradičný prístup si vyžaduje manuálne úpravy pri každej novej implementácii platformy: migrácia údajov, zmena názvov atribútov, úprava formátov. Tieto kroky nezlepšujú kvalitu samotných údajov, ale len maskujú problémy, čím vytvárajú začarovaný kruh nekonečných transformácií. V dôsledku toho sa spoločnosti stávajú závislé od konkrétnych softvérových riešení a digitálna transformácia sa spomaľuje.

V nasledujúcich kapitolách sa pozrieme na to, ako správne štruktúrovať údaje a ako potom vytvoriť univerzálné modely, minimalizovať závislosť od platformy a zamerať sa na to najdôležitejšie - údaje ako strategický zdroj, okolo ktorého sa budujú udržateľné procesy.

#### Naučte sa, ako meniť dokumenty, súbory PDF, obrázky a texty do štruktúrovaných formátov.

V stavebných projektoch existuje prevažná väčšina informácií v neštruktúrovanej forme: technické dokumenty, výkazy prác, výkresy, špecifikácie, harmonogramy, protokoly. Ich rôznorodosť - čo do formátu aj obsahu - sťažuje integráciu a automatizáciu.

Proces konverzie do štruktúrovaných alebo pološtruktúrovaných formátov sa môže lísiť v závislosti od typu vstupných údajov a požadovaných výsledkov spracovania.

Transformácia údajov z neštruktúrovanej do štruktúrovanej podoby je umenie aj veda. Tento proces sa lísi v závislosti od typu vstupných údajov a účelu analýzy a často zaberá značnú časť práce dátového inžiniera (obrázok 3.2-5) a analytika, pričom cieľom je vytvoriť čistý, usporiadany súbor údajov.



Obr. 4.1-1 Konverzia neštruktúrovaného naskenovaného dokumentu do formátu štruktúrovanej tabuľky.

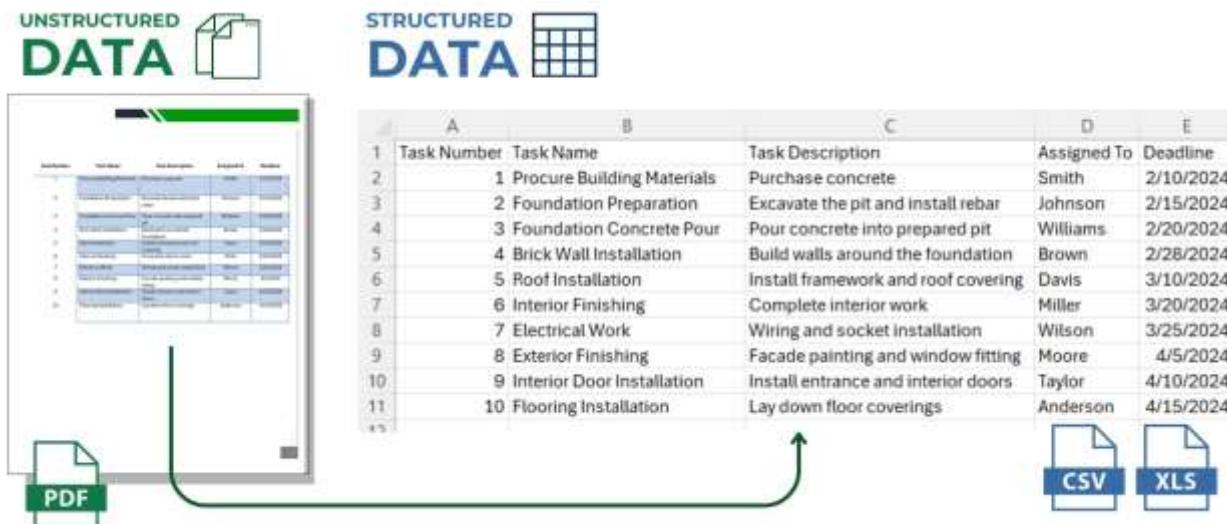
Prevedenie dokumentov, PDF, obrázkov a textov do štruktúrovaného formátu (obrázok 4.1-1) je postupný proces, ktorý zahrňa nasledujúce kroky:

- **Výpis:** V tomto kroku sa načíta zdrojový dokument alebo obrázok obsahujúci neštruktúrované údaje. Môže to byť napríklad PDF -dokument, fotografia, výkres alebo schéma.
  - **Transformácia údajov** (Transform): Nasleduje krok transformácie neštruktúrovaných údajov do štruktúrovaného formátu. Môže ísť napríklad o rozpoznávanie a interpretáciu textu z obrázkov pomocou optického rozpoznávania znakov (OCR) alebo iných metód spracovania.
  - **Načítanie a uloženie údajov** (Load): posledný krok zahŕňa uloženie spracovaných údajov v rôznych formátoch, ako sú CSV, XLSX, XML, JSON, na ďalšiu prácu, pričom výber formátu závisí od konkrétnych požiadaviek a preferencií.

Tento proces, známy ako ETL (Extract, Transform, Load), zohráva kľúčovú úlohu pri automatizovanom spracovaní údajov a podrobnejšie sa ním budeme zaoberať v kapitole "ETL a Pipeline: Extract, Transform, Load". Ďalej sa budeme venovať príkladom transformácie dokumentov rôznych formátov na štruktúrované údaje.

Príklad konverzie PDF -dokumentu na tabuľku

Jednou z najčastejších úloh v stavebných projektoch je spracovanie špecifikácií vo formáte PDF. Na demonštráciu prechodu z neštruktúrovaných údajov do štruktúrovaného formátu si uvedme praktický príklad: extrahovanie tabuľky z dokumentu PDF a jej prevod do formátu CSV alebo Excel (obr. 4.1-2).



Obr. 4.1-2 Na rozdiel od PDF, formáty CSV a XLSX sú rozšírené a ľahko sa integrujú do rôznych systémov na správu údajov.

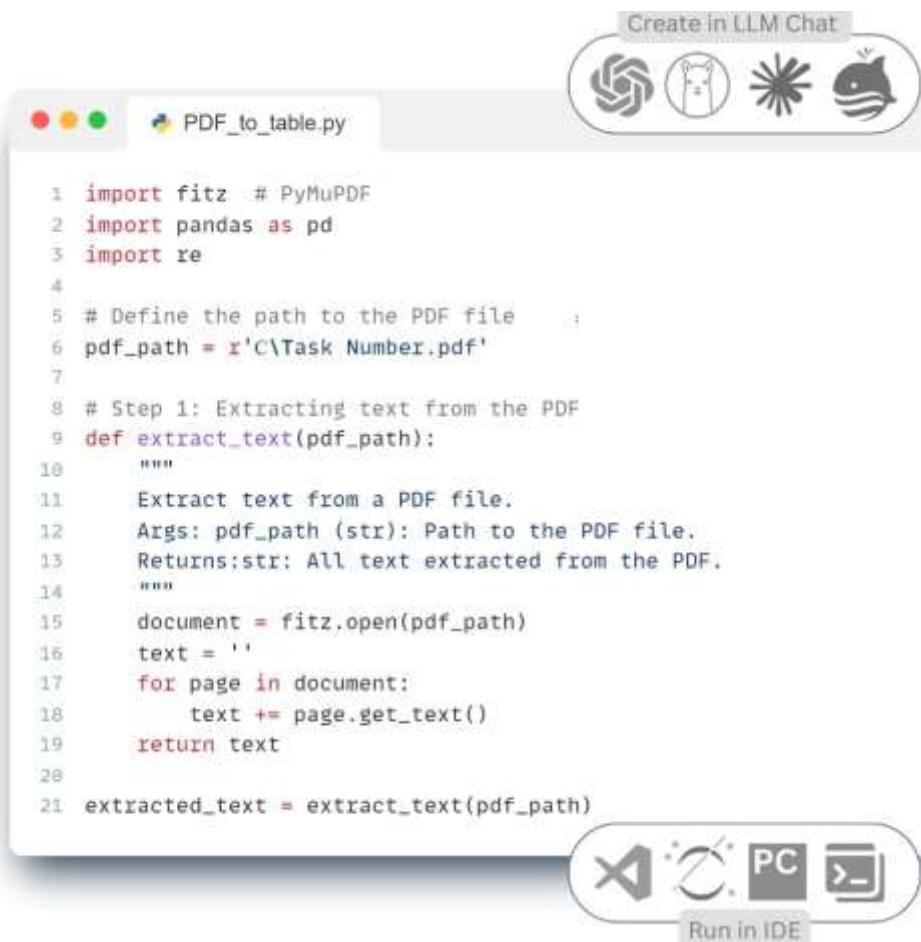
Jazykové modely LLM, ako napríklad ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN výrazne zjednodušujú spôsob, akým dátoví vedci pracujú s údajmi, znižujú potrebu hlbokého učenia programovacích jazykov a umožňujú riešiť mnohé úlohy pomocou textových dotazov.

Preto namiesto toho, aby sme trávili čas hľadaním riešení na internete (zvyčajne na webovej stránke StackOverflow alebo tematických fórch a chatoch) alebo kontaktovaním odborníkov na spracovanie údajov, môžeme využiť možnosti moderných online alebo lokálnych LLM. Stačí položiť dotaz a model poskytne hotový kód na konverziu PDF -dokumentu do tabuľkového formátu.

- ❷ Pošlite nasledujúcu textovú požiadavku na ľubovoľný model LLM (CHATGRT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akýkoľvek iný):

Napíšte prosím kód na extrakciu textu zo súboru PDF, ktorý obsahuje tabuľku. Kód by mal prijať cestu k súboru ako argument a vrátiť vyextrahovanú tabuľku ako DataFrame ↵

- ❷ Odpoveď LLM -modelu bude vo väčšine prípadov vo forme kódu v jazyku Python , keďže tento jazyk sa široko používa na spracovanie údajov, automatizáciu a prácu s rôznymi formátkami súborov:



The screenshot shows a window titled "PDF\_to\_table.py". The code is as follows:

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11         Extract text from a PDF file.
12         Args: pdf_path (str): Path to the PDF file.
13         Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

At the top right, there is a "Create in LLM Chat" button and a row of icons for different AI models. At the bottom right, there is a "Run in IDE" button.

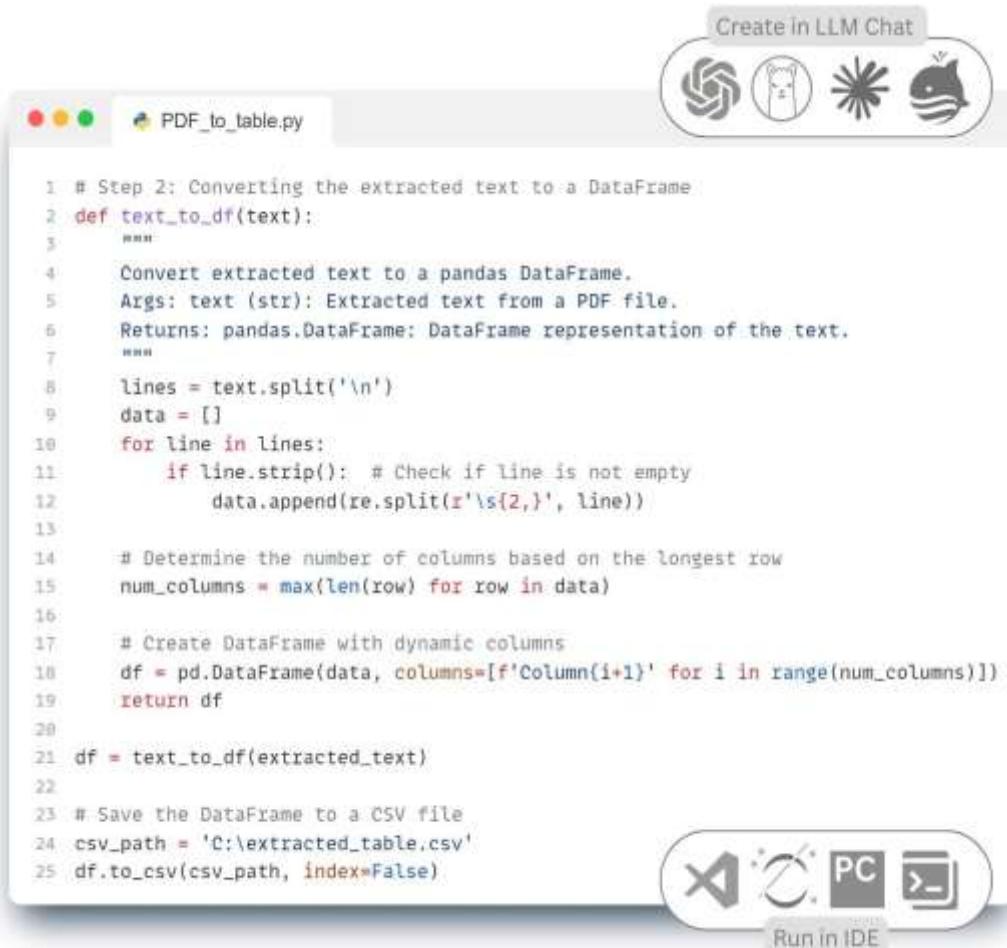
Obr. 4.1-3 Odpoveď LLM vo forme kódu jazyka Python a jeho knižníc a balíkov (Pandas, Fitz) extrahuje text z PDF - súboru.

Tento kód (obrázok 4.1-3) možno spustiť offline v jednom z populárnych IDE, ktoré sme spomenuli vyššie: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo populárne online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- ❸ V kroku "Convert" použijeme populárnu knižnicu Pandas (ktorej sme sa podrobne venovali v kapitole "Python Pandas: nepostrádateľný nástroj na prácu s údajmi") na načítanie extrahovaného textu do DataFrame a uloženie DataFrame do súboru tabuľky CSV alebo XLXS:

Potrebujem kód, ktorý prevedie výslednú tabuľku zo súboru PDF -file do DataFrame.  
Tiež pridať kód na uloženie DataFrame do súboru CSV ↵

💡 Odpoveď LLM :



The screenshot shows a code editor window titled "PDF\_to\_table.py". The code is as follows:

```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

At the top right of the code editor are four circular icons representing different AI models. At the bottom right are three more circular icons representing different environments or tools.

Obr. 4.1-4 Konverzia extrahovanej tabuľky z PDF do DataFrame a uloženie tabuľky do CSV - súboru.

Ak sa počas vykonávania kódu vyskytne chyba (obr. 4.1-3, obr. 4.1-4) - napr. v dôsledku chýbajúcich knižníc alebo nesprávnej cesty k súboru - text chyby možno jednoducho skopírovať spolu so zdrojovým kódom a znova odoslať do LLM -modelu. Model analyzuje chybovú správu, vysvetlí, v čom je problém, a navrhne opravy alebo ďalšie kroky.

Takto sa interakcia s AI LLM stáva kompletným cyklo→→test→ spätná väzba→ korekcia  
- bez potreby hlbokých technických znalostí.

Pomocou obyčajného textového dotazu v LLM chat a desiatok riadkov jazyka Python, ktorý môžeme spustiť lokálne v ľubovoľnom IDE, sme konvertovali dokument PDF -dokument na tabuľkový formát CSV, ktorý je na rozdiel od dokumentu PDF ľahko strojovo čitateľný a rýchlo integrovateľný do akéhokoľvek systému na správu údajov.

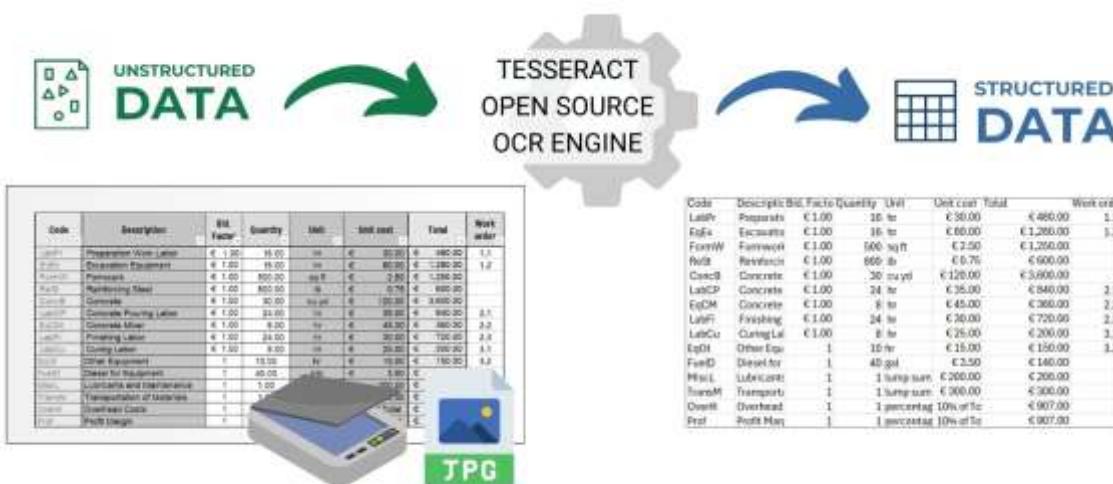
Tento kód (obrázok 4.1-3, obrázok 4.1-4) môžeme skopírovaním z ľubovoľného chatu LLM použiť na desiatky alebo tisíce nových dokumentov PDF na serveri, čím sa automatizuje proces konverzie prúdu neštruktúrovaných dokumentov do štruktúrovaného formátu tabuľky CSV.

Dokumenty PDF však nie vždy obsahujú text, častejšie ide o naskenované dokumenty, ktoré je potrebné spracovať ako obrázky. Hoci sú obrázky vo svojej podstate neštruktúrované, vývoj a aplikácia knižníc rozpoznávania nám umožňuje extrahovať, spracovať a analyzovať ich obsah, čo nám umožňuje plne využívať tieto údaje v našich obchodných procesoch.

## Konvertovať obrázok JPEG, PNG do štruktúrovanej podoby

Obrázky sú jednou z najbežnejších foriem neštruktúrovaných údajov. V stavebníctve a mnohých ďalších odvetviach sa obrovské množstvo informácií ukladá vo forme naskenovaných dokumentov, schém, fotografií a výkresov. Takéto údaje obsahujú cenné informácie, ale nedajú sa priamo spracovať, napríklad v tabuľkovom procesore Excel alebo v databáze. Obrázky obsahujú množstvo komplexných informácií, pretože ich obsah, farby, textúry sú rôznorodé a na získanie užitočných informácií je potrebné špeciálne spracovanie.

Problémom pri používaní obrázkov ako zdroja údajov je nedostatočná štruktúra. Obrázky neprenášajú význam priamym, ľahko kvantifikateľným spôsobom, ktorý by počítač mohol okamžite pochopiť alebo spracovať, ako to robí tabuľka v programe Excel alebo tabuľka v databáze. Na prevod neštruktúrovaných obrazových údajov do štruktúrovanej podoby sa musia použiť špeciálne knižnice schopné interpretovať vizuálne informácie, ktoré obsahujú (obrázok 4.1-5).



Obr. 4.1-5 Konverzia naskenovaných dokumentov a obrázkov do štruktúrovanej podoby je možná pomocou špeciálnych nástrojov OCR.

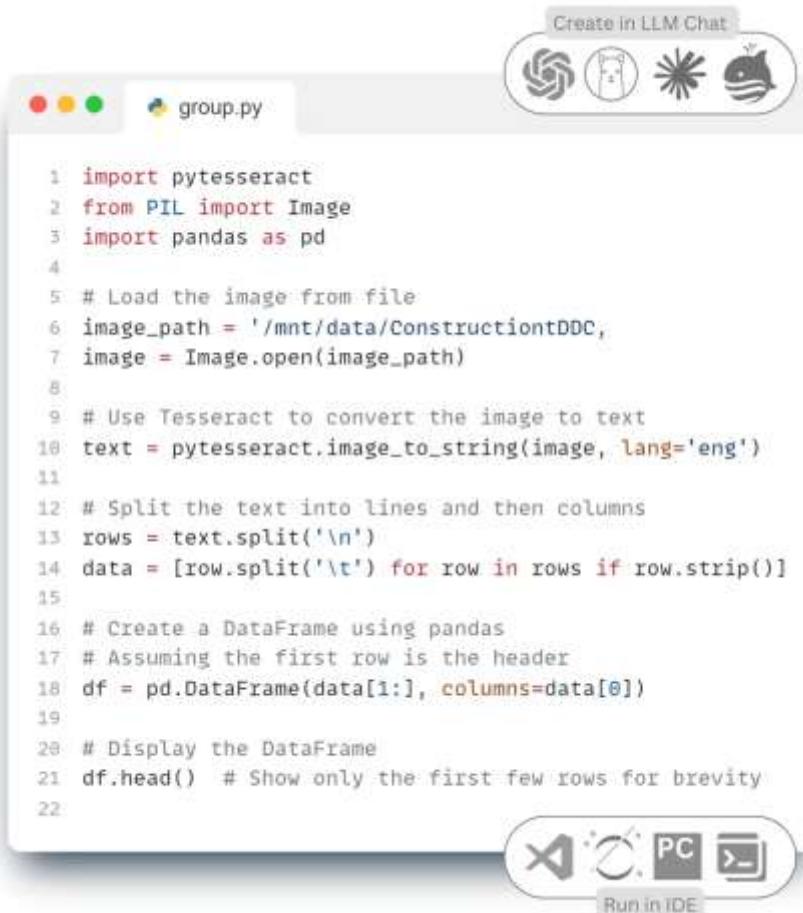
Technológia OCR (Optické rozpoznávanie znakov) sa používa na extrakciu textu z obrázkov. Umožňuje rozpoznať písmená a číslice v skenovaných dokumentoch, fotografiách a súboroch PDF a premieňať ich na upraviteľný a strojovo čitateľný text. Technológie OCR sa už dlho používajú pri automatizácii dokumentov a dnes sa ľahko integrujú do akýchkoľvek obchodných procesov a aplikácií Python. Jedným z najpopulárnejších nástrojov OCR je Tesseract, open source, pôvodne vyvinutý spoločnosťou HP™ a teraz podporovaný spoločnosťou Google™. Podporuje viac ako 100 jazykov a má vysokú presnosť rozpoznávania.

Požiadajme LLM chat, aby napísal vzorový kód na získanie údajov z naskenovanej alebo odfotografovanej tabuľky štruktúrovaným spôsobom.

- ❷ Pošlite textovú požiadavku na chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akýkoľvek iný):

Napíšte kód na konverziu obrázka JPEG obsahujúceho tabuľku do tabuľky DataFrame ↴

- 2 V odpovedi LLM sa vo väčšine prípadov navrhuje použitie knižnice Pytesseract na rozpoznávanie textu v obrázkoch :



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC'
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

Obr. 4.1-6 Konvertovanie textu získaného z tabuľky obrázkov alebo fotografií do štruktúrovanej tabuľkovej reprezentácie.

V tomto príklade - kód (obrázok 4.1-6) odvodený z LLM, používa knižnicu pytesseract (Tesseract pre Python) na prevod obrázka na text pomocou OCR (optické rozpoznávanie znakov) a knižnicu Pandas na prevod tohto textu do štruktúrovanej formy, t. j. DataFrame.

Proces konverzie zvyčajne zahŕňa predbežné spracovanie na zlepšenie kvality obrazu, po ktorom sa použijú rôzne algoritmy na detekciu vzorov, extrakciu prvkov alebo rozpoznávanie objektov. Výsledkom je konverzia neštruktúrovaných vizuálnych informácií na štruktúrované údaje.

Hoci kľúčovými zdrojmi neštruktúrovaných informácií sú súbory PDF a obrázky, skutočným šampiónom z hľadiska objemu je text generovaný v e-mailoch, chatoch, stretnutiach, messengeroch. Týchto údajov je nielen veľa - sú rozptýlené, neformalizované a mimoriadne zle štruktúrované.

## Konverzia textových údajov do štruktúrovanej podoby

Okrem dokumentov vo formáte PDF s tabuľkami (obr. 4.1-2) a naskenovaných verzií tabuľkových formulárov (obr. 4.1-5) sa značná časť informácií v projektovej dokumentácii uvádza v textovej podobe. Môže ísť o súvislé vety v textových dokumentoch aj o útržkovité záznamy roztrúsené po výkresoch a schémach. V podmienkach moderného spracovania údajov je jednou z najčastejších úloh prevod takéhoto textu do štruktúrovanej podoby vhodnej na analýzu, vizualizáciu a rozhodovanie.

Ústredným prvkom tohto procesu je taxonómia, klasifikačný systém, ktorý umožňuje organizovať informácie do kategórií a podkategórií na základe spoločných znakov.

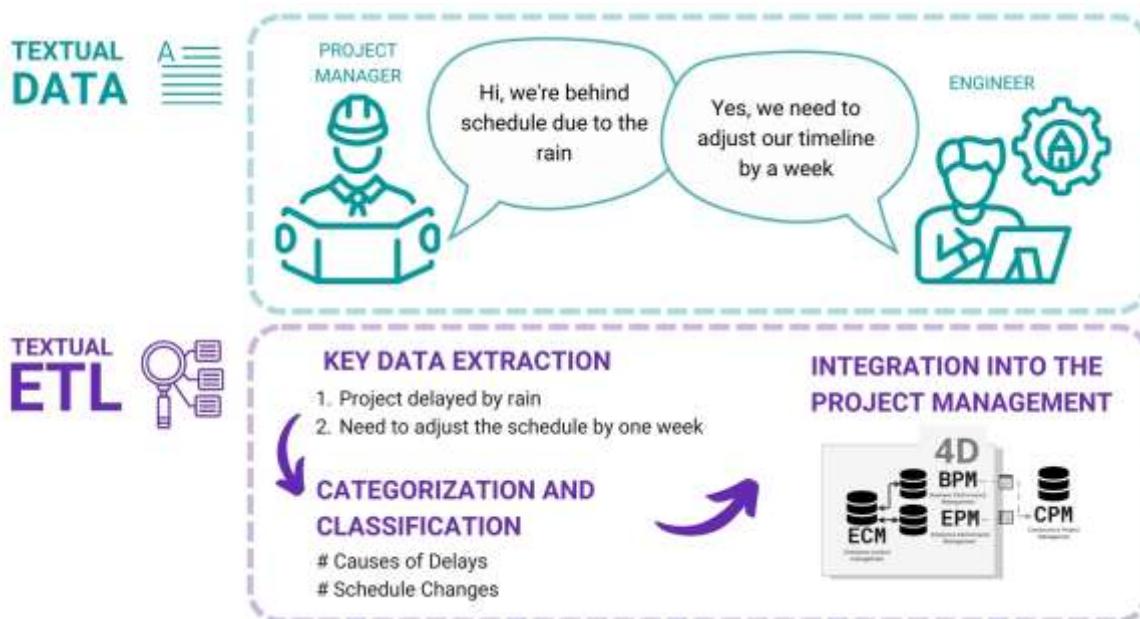
**Taxonómia** je hierarchická klasifikačná štruktúra používaná na zoskupovanie a organizáciu objektov. V kontexte spracovania textu slúži ako základ na systematické kategorizovanie položiek do sémantických kategórií, čím sa zjednodušuje analýza a zvyšuje kvalita spracovania údajov.

Vytváranie taxonómie sprevádzajú kroky extrakcie entít, kategorizácie a kontextualizácie. Na modelovanie procesu extrakcie informácií z textových údajov slúžia nasledujúce kroky podobné tým, ktoré sme už použili pri štruktúrovaní údajov z dokumentov PDF:

- **Výpis**): potrebujete analyzovať textové údaje, aby ste získali informácie o oneskoreniach a zmenách v harmonograme projektu.
- **Kategorizácia a klasifikácia** (Transofrm): kategorizácia prijatých informácií, napr. dôvody meškania a zmeny harmonogramu.
- **Integrácia** (Load): na záver pripravíme štruktúrované údaje na integráciu do externých systémov správy údajov.

Zoberme si situáciu: máme dialóg medzi projektovým manažérom a inžinierom, ktorý diskutuje o problémoch s harmonogramom. Naším cieľom je vybrať kľúčové prvky (dôvody meškania, úpravy harmonogramu) a prezentovať ich štruktúrovaným spôsobom (obrázok 4.1-7).

Vykonajme extrakciu na základe očakávaných kľúčových slov, vytvorime DataFrame na simuláciu extrakcie údajov a po transformácii novú tabuľku DataFrame, ktorá bude obsahovať stĺpce pre dátum, udalosť (napr. dôvod meškania) a akciu (napr. zmena plánu).



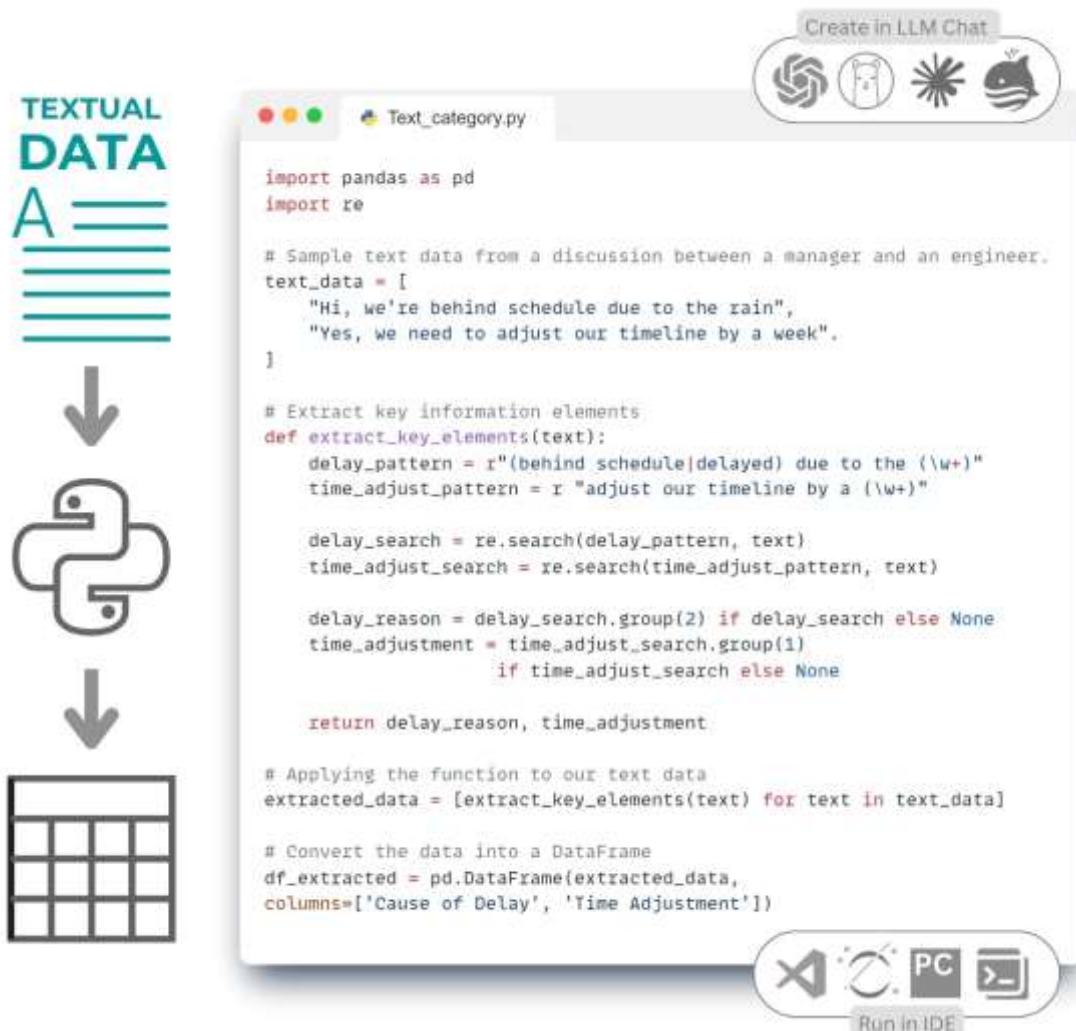
Obrázok 4.1-7 Zvýraznenie kľúčových informácií z textu o potrebe úpravy časového harmonogramu a integrácie zmien do systému riadenia projektu.

Tu je kód na riešenie problému pomocou textového dotazu v jednom z jazykových modelov ako v predchádzajúcich príkladoch.

❷ Pošlite textovú požiadavku do ktorejkoľvek chatovej miestnosti LLM:

Vediem rozhovor medzi manažérom: "Dobrý deň, kvôli dažďu meškáme s termínom" a inžinierom: "Áno, potrebujeme upraviť termín o týždeň". Potrebujem skript, ktorý bude analyzovať budúce podobné textové dialógy, získa z nich dôvody meškania a potrebných úprav termínov a potom z týchto údajov vygeneruje DataFrame. DataFrame by sa potom mal uložiť do súboru CSV. ↵

- Odpoveď z LLM bude zvyčajne obsahovať kód jazyka Python -kód využívajúci regulárne výrazy (re - Regex) a knižnicu Pandas (pd) :

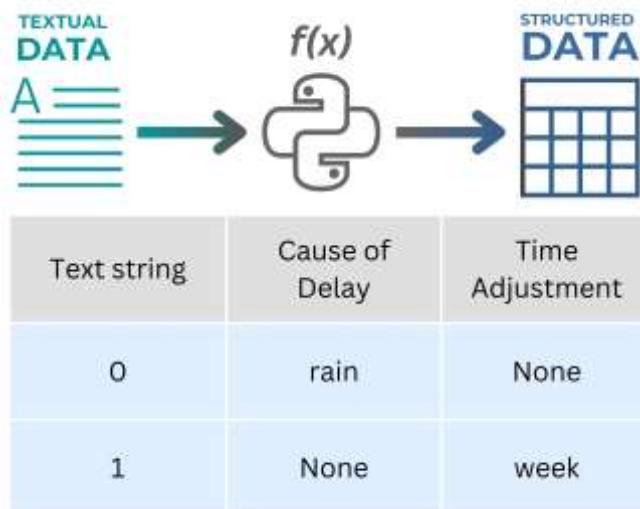


Obrázok 4.1-8 Zvýraznenie klúčových informácií z textu o potrebe úpravy termínov v tabuľke.

V tomto príklade (obrázok 4.1-7) sa analyzujú textové údaje obsahujúce korešpondenciu medzi projektovým manažérom a inžinierom s cieľom identifikovať a získať špecifické informácie, ktoré môžu ovplyvniť riadenie budúcich projektov s podobnými dialógmi. Pomocou regulárnych výrazov (viac o regulárnych výrazoch si povieme v kapitole "Štruktúrované požiadavky a regulárne výrazy RegEx") sa prostredníctvom vzorov identifikujú príčiny meškania projektu a potrebné úpravy časového harmonogramu. Funkcia napísaná v tomto príklade vyberie z reťazcov bud' príčinu meškania, alebo úpravu času na základe vzorov: vyberie slovo za "kvôli" ako príčinu meškania alebo slovo za "podľa" ako úpravu času.

Ak sa v riadku spomína meškanie spôsobené počasím, ako príčina sa identifikuje "dážd"; ak sa v riadku

spomína úprava harmonogramu pre určité obdobie, toto obdobie sa vyčlení ako časová úprava (obrázok 4.1-9). Neprítomnosť niektorého z týchto slov v riadku má za následok hodnotu "Žiadne" pre príslušný stĺpec atribútov.



Obr. 4.1-9 Súhrnná tabuľka získaná ako DataFrame, po vykonaní kódu, obsahuje informácie o existencii oneskorení a potrebných časových úpravách.

Štruktúrovanie a parametrizácia podmienok z textu (dialógu, listu, dokumentu) umožňuje rýchlo eliminovať meškanie pri výstavbe: napríklad nedostatok pracovníkov môže ovplyvniť tempo prác pri zlom počasí, preto firmy, ktoré poznajú parametre meškania z dialógov (obr. 4.1-9) medzi majstrom na stavbe a projektovým manažérom - vopred môžu posilniť posádku v prípade nepriaznivej predpovede.

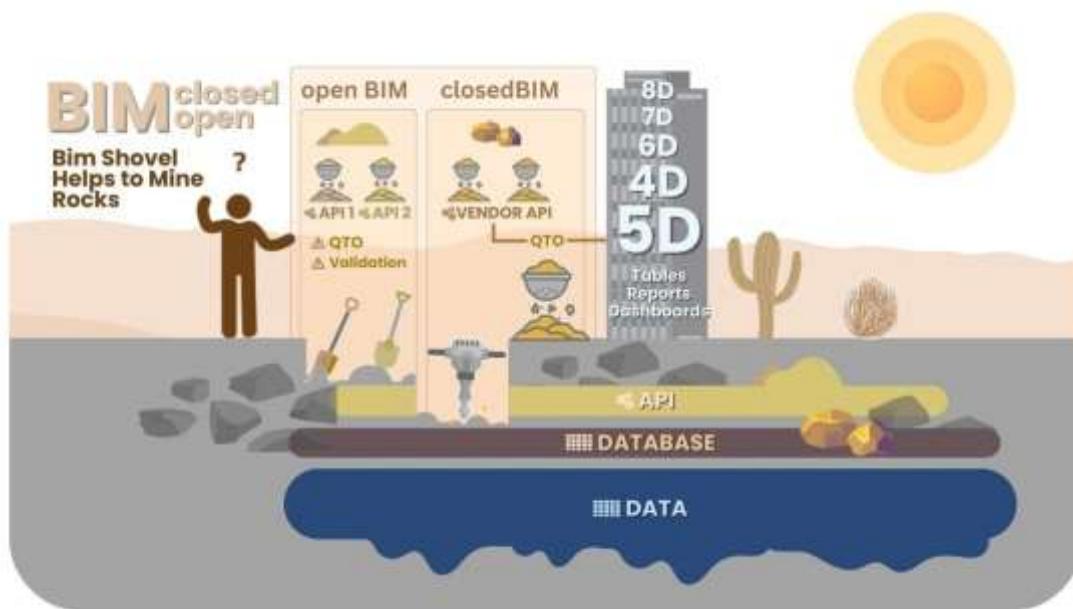
Konverziu dokumentov a obrázkov do štruktúrovaného formátu možno dosiahnuť pomocou pomerne jednoduchých, otvorených a bezplatných nástrojov založených na kategorizácii.

Kategorizácia prvkov je tiež klúčovou súčasťou práce s údajmi o projekte, najmä v súvislosti s používaním softvéru CAD (BIM).

## Konverzia údajov CAD (BIM) do štruktúrovanej podoby

Štruktúrovanie a kategorizácia údajov CAD (BIM) je zložitejšia úloha, pretože údaje uložené z databáz CAD (BIM) sú takmer vždy v uzavretých alebo zložitých parametrických formátoch a často kombinujú geometrické dátové prvky (pološtruktúrované) a metainformačné prvky (pološtruktúrované alebo štruktúrované údaje) súčasne.

Natívne formáty údajov v systémoch CAD (BIM) sú zvyčajne chránené a neprístupné na priame použitie, pokiaľ nie je k dispozícii špecializovaný softvér alebo API - rozhrania samotného vývojára (obr. 4.1-10). Takáto izolácia údajov vytvára uzavreté úložné silá, ktoré obmedzujú voľnú výmenu informácií a bránia vytváraniu komplexných digitálnych procesov v podniku.



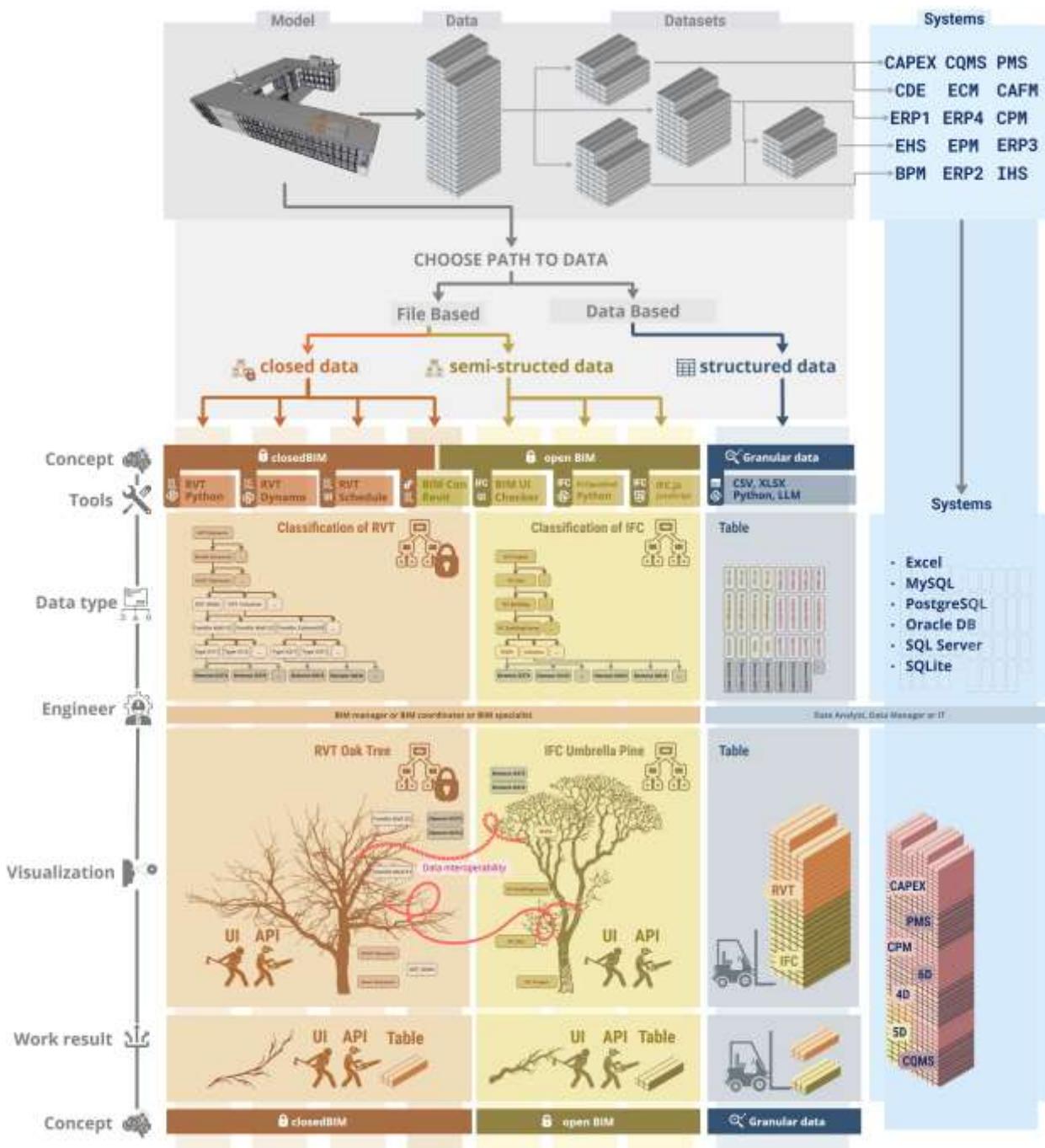
Obrázok 4.1-10 Profesionáli CAD (BIM) môžu pristupovať k natívnym údajom prostredníctvom API -prípojok alebo nástrojov dodávateľa.

V špeciálnych formátoch CAD (BIM) sa informácie o vlastnostiach a atribútoch prvkov projektu zhromažďujú v hierarchickom klasifikačnom systéme, kde sa entity s príslušnými vlastnosťami nachádzajú, podobne ako plody ovocného stromu, v najnovších uzloch vetiev klasifikácie údajov (obr. 4.1-11).

Extrakcia údajov z takýchto hierarchií je možná dvoma spôsobmi: buď manuálne, kliknutím na každý uzol, akoby ste spracovávali strom a sekerou osekávali vybrané vetvy kategórií a typov. Alternatívou je použitie aplikačných programových rozhraní (API), ktoré poskytujú efektívnejší, automatizovaný prístup k získavaniu a zoskupovaniu údajov, prípadne ich transformácií do štruktúrovanej tabuľky na použitie v iných systémoch.

Na extrakciu štruktúrovaných dátových tabuľiek z projektov CAD (BIM) možno použiť rôzne nástroje, ako napríklad Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC alebo riešenia s otvoreným zdrojovým kódom, napríklad IfcOpSh alebo IFCjs pre formát IFC.

Moderné nástroje na export a konverziu údajov umožňujú zjednodušiť spracovanie a prípravu údajov rozdelením obsahu CAD modelov na dve kľúčové zložky: geometrické informácie a atribútové údaje (obr. 4.1-13) - metainformácie popisujúce vlastnosti konštrukčných prvkov (obr. 3.1-16). Tieto dve vrstvy údajov zostávajú prepojené prostredníctvom jedinečných identifikátorov, vďaka ktorým je možné presne priradiť každý prvek s popisom geometrie (prostredníctvom parametrov alebo polígonov) k jeho atribútom: názvu, materiálu, stupňu dokončenia, nákladom atď. Tento prístup zabezpečuje integritu modelu a umožňuje flexibilné využívanie údajov na vizualizáciu (údaje geometrického modelu) aj na analytické alebo riadiace úlohy (štruktúrované alebo voľne štruktúrované), pričom s týmito dvoma typmi údajov možno pracovať oddelene alebo súbežne.



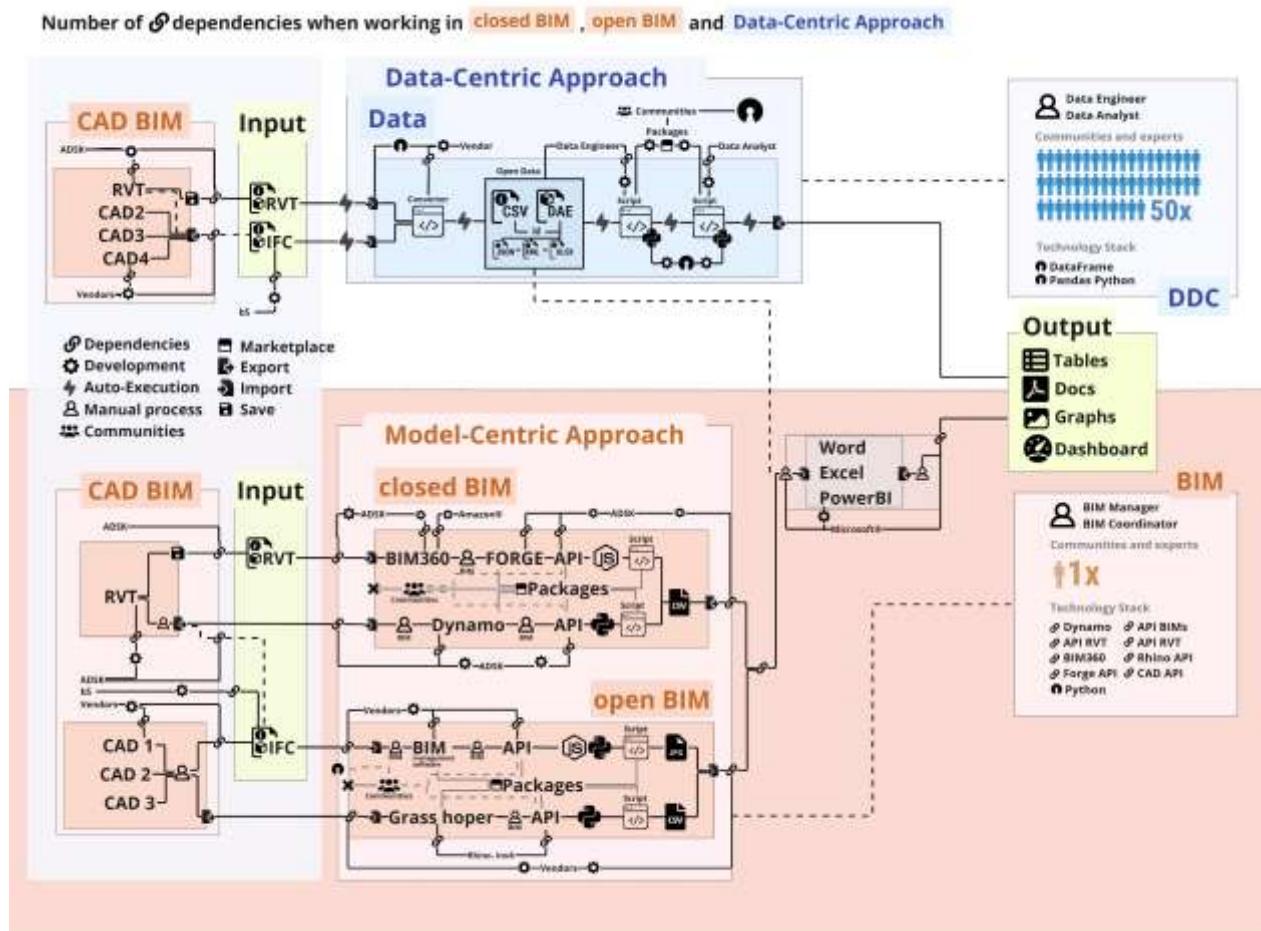
Obr. 4.1-11 Zobrazenie informácií z databáz CAD (BIM) sa používateľovi prezentuje vo forme klasifikačných stromov.

S rozvojom technológií reverzného inžinierstva a príchodom SDK (Software Development Kit) na konverziu údajov CAD sa dostupnosť a konverzia údajov z uzavretých formátov softvéru CAD (BIM) stala oveľa jednoduchšou. V súčasnosti je možné legálne a bezpečne konvertovať údaje z uzavretých formátov do univerzálnych formátov vhodných na analýzu a použitie v iných systémoch. História prvých nástrojov na reverzné inžinierstvo ("Open DWG") a boj o nadvládu nad formátmi dodávateľov

CAD bola rozobraná v kapitole "Štruktúrované údaje: základ digitálnej transformácie".

Nástroje na reverzné inžinierstvo umožňujú legítimne získavanie údajov z uzavretých proprietárnych formátov, rozkladajú informácie zo zmiešaného formátu CAD (BIM) na typy údajov a formáty požadované používateľom, čím uľahčujú ich spracovanie a analýzu.

Pomocou reverzného inžinierstva a priameho prístupu k informáciám z databáz CAD sprístupňuje informácie, umožňuje otvorené údaje a otvorené nástroje, ako aj analýzu údajov pomocou štandardných nástrojov, vytváranie prehľadov, vizualizácií a integráciu s inými digitálnymi systémami (obr. 4.1-12).



Obr. 4.1-12 Priamy prístup k údajom CAD minimalizuje závislosť od softvérových platform a smeruje k dátovo orientovanému prístupu.

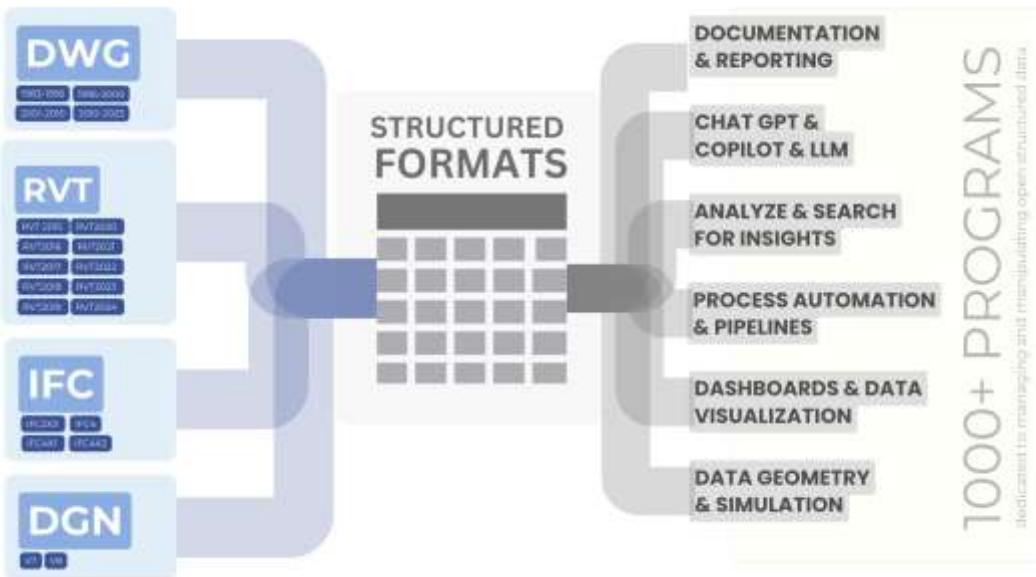
Od roku 1996 pre formát DWG, od roku 2008 pre formát DGN a od roku 2018 pre formát RVT je možné pomocou nástrojov reverzného inžinierstva pohodlne a efektívne konvertovať pôvodne uzavreté formáty CAD do ľubovoľných iných formátov vrátane štruktúrovaných formátov (obr. 4.1-13). V súčasnosti takmer všetky významné CAD (BIM) a veľké strojárske spoločnosti na svete používajú SDK - nástroje reverzného inžinierstva na extrakciu údajov z uzavretých formátov CAD (BIM) dodávateľa [92].



Obr. 4.1-13 Použitie nástrojov reverzného inžinierstva umožňuje konvertovať databázy programov CAD (BIM) na akýkoľvek vhodný dátový model.

Konverzia údajov z uzavretých proprietárnych formátov do otvorených formátov a rozdelenie zmiešaných formátov CAD (BIM) na geometrické a metainformačné atribútové údaje zjednodušuje proces práce s nimi, sprístupňuje ich na analýzu, manipuláciu a integráciu s inými systémami (obr. 4.1-14).

Pri dnešnej práci s údajmi CAD (BIM) sme sa dostali do bodu, keď na prístup k informáciám z formátov CAD (BIM) nemusíme žiadať povolenie od dodávateľov CAD (BIM).



Obr. 4.1-14 Moderné nástroje SDK umožňujú legálnu konverziu údajov z proprietárnych formátov databáz CAD (BIM).

Súčasné trendy v spracovaní údajov CAD dizajnu naďalej formujú kľúčoví hráči na trhu - dodávateelia CAD, ktorí sa snažia posilniť svoju pozíciu vo svete údajov a vytvárajú nové formáty a koncepcie.

## Dodávateľia riešení CAD prechádzajú na štruktúrované údaje

Od roku 2024 prechádza odvetvie projektovania a výstavby významnou technologickou zmenou v oblasti využívania a spracovania údajov. Namiesto voľného prístupu k údajom o projektovaní sa dodávateľia CAD -systémov zameriavajú na podporu ďalších nových konceptov. Prístupy ako BIM (vytvorený v roku 2002) a otvorený BIM (vytvorený v roku 2012) postupne ustupujú moderným technologickým riešeniam, ktoré začínajú presadzovať dodávateľia CAD [93]:

- Prechod na používanie "granulovaných" údajov, ktoré umožňujú efektívne riadenie informácií a prechod na analýzu údajov
- Vznik formátu USD a implementácia prístupu Entity-component-system (ECS) na flexibilnú organizáciu údajov
- Aktívne využívanie umelej inteligencie pri spracovaní údajov, automatizácii procesov a analýze údajov
- Vývoj interoperability - zlepšenie interakcie medzi rôznymi programami, systémami a databázami

Každý z týchto aspektov bude podrobnejšie rozobraný v šiestej časti knihy "CAD a BIM: marketing, realita a budúcnosť projektových údajov v stavebnictve". V rámci tejto kapitoly len stručne načrtнемe všeobecný vektor zmien: hlavní dodávateľia CAD sa v súčasnosti snažia prehodnotiť spôsob štruktúrovania informácií o projektovaní. Jedným z kľúčových posunov je opustenie klasického súborového modelu ukladania dát v prospech granulárnej, analyticky orientovanej dátovej architektúry, ktorá poskytuje nepretržitý prístup k jednotlivým komponentom modelu [93].

Podstatou tohto procesu je, že priemysel postupne upúšťa od ťažkopádnych, špecializovaných a parametrických formátov, ktoré si vyžadujú geometrické jadrá, v prospech univerzálniejsích, strojovo čitateľných a flexibilnejších riešení.

Jedným z takýchto hnacích motorov zmien je formát USD (Universal Scene Description), ktorý bol pôvodne vyvinutý v odvetví počítačovej grafiky, ale vďaka vývoju platformy NVIDIA Omniverse (a Isaac Sim) na simulácie a vizualizácie [93] je už uznávaný v technických aplikáciách. Na rozdiel od parametrického IFC, USD ponúka jednoduchšiu štruktúru a umožňuje opis geometrie a vlastností objektov vo formáte JSON (obrázok 4.1-15), čo uľahčuje spracovanie informácií a urýchľuje ich integráciu do digitálnych procesov. Nový formát umožňuje ukladať geometriu (okrem BREP -NURBS - podrobnejšie v 6. časti knihy) vo forme polygónov MESH a vlastnosti objektov v JSON, čo je výhodnejšie pre automatizované procesy a prácu v clouдовých ekosystémoch [94].

Niektorí dodávateľia CAD a ERP už používajú podobné formáty (napr. NWD, SVF, CP2, CPIXML), ale väčšina z nich zostáva uzavretá a nedostupná pre externé použitie, čo obmedzuje možnosti integrácie a opakovaného použitia údajov. V tejto súvislosti môže USD zohrávať rovnakú úlohu ako svojho času DXF - otvorenú alternatívu k proprietárny formátom, ako je DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

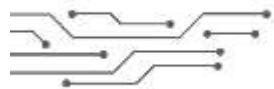
Obr. 4.1-15 USD formát ako pokus dodávateľov CAD o splnenie požiadavky na interoperabilitu a nezávislosť návrhových údajov od geometrických jadier.

Prechod hlavných vývojárov na otvorené a zjednodušené formáty USD, GLTF, OBJ, XML (uzavreté formáty NWD, CP2, SVF, SVF2, CPIXML) a podobné formáty (obr. 3.1-17) odráža celosvetový trend a dopyt priemyslu po zjednodušení údajov a ich väčšej dostupnosti. V nasledujúcich rokoch možno očakávať postupný odklon od zložitých parametrických noriem a formátov so závislosťou od geometrických jadier v prospech ľahších a štruktúrovanejších riešení. Tento prechod urýchli digitalizáciu stavebníctva, uľahčí automatizáciu procesov a zjednoduší výmenu údajov.

Napriek strategickým plánom dodávateľov CAD na podporu nových otvorených formátov môžu odborníci v stavebníctve získať plný prístup k údajom z uzavretých systémov CAD aj bez potreby nástrojov CAD (BIM), a to pomocou nástrojov reverzného inžinierstva.

Všetky tieto trendy nevyhnutne vedú k prechodu od objemných, monolitických 3D modelov k univerzálnym, štruktúrovaným údajom a k používaniu formátov, ktoré sa už dávno osvedčili v iných odvetviach. Akonáhle projektové tímy začnú vnímať modely CAD nielen ako vizuálne objekty alebo súbor súborov, ale ako databázy obsahujúce znalosti a informácie, prístup k projektovaniu a správe sa dramaticky zmení.

Ked' sa tímy naučia získať štruktúrované údaje z dokumentov, textov, výkresov a modelov CAD a majú prístup k databázam, ďalším klúčovým krokom je modelovanie údajov a zabezpečenie kvality. Práve tento krok do značnej miery určuje rýchlosť spracovania a transformácie informácií, ktoré sa v konečnom dôsledku použijú na rozhodovanie v konkrétnych aplikačných úlohách.



## KAPITOLA 4.2.

### **KLASIFIKÁCIA A INTEGRÁCIA: SPOLOČNÝ JAZYK PRE STAVEBNÉ ÚDAJE**

#### **Rýchlosť rozhodovania závisí od kvality údajov**

Dnešná architektúra údajov o dizajne prechádza zásadnými zmenami. Odvetvie prechádza od objemných, izolovaných modelov a uzavretých formátov k flexibilnejším, strojovo čitateľným štruktúram zameraným na analýzu, integráciu a automatizáciu procesov. Samotný prechod na nové formáty však nezaručuje efektívnosť - v centre pozornosti je nevyhnutne kvalita samotných údajov.

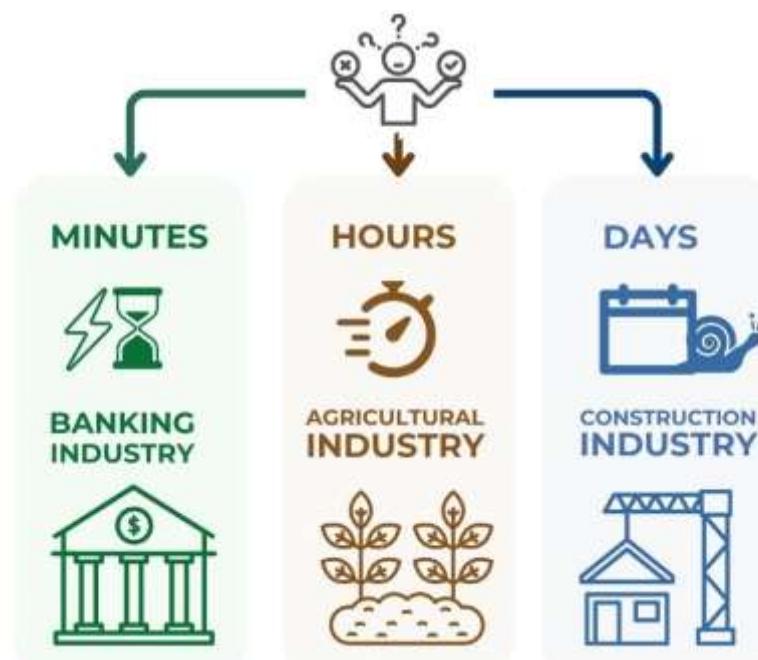
Na stránkach tejto knihy veľa hovoríme o formátoch, systémoch a procesoch. Všetky tieto snahy však nemajú význam bez jedného klíčového prvku: údajov, ktorým možno dôverovať. Kvalita údajov je základným kameňom digitalizácie a budeme sa k nej vracať v nasledujúcich častiach.

Moderné stavebné firmy - najmä tie veľké - používajú desiatky a niekedy aj tisíce rôznych systémov a databáz (obr. 4.2-1). Tieto systémy sa musia nielen pravidelne napĺňať novými informáciami, ale aj navzájom efektívne spolupracovať. Všetky nové údaje, ktoré vznikajú ako výsledok spracovania prichádzajúcich informácií, sa integrujú do týchto prostredí a slúžia na riešenie konkrétnych obchodných úloh.

A zatiaľ čo v minulosti rozhodovali o konkrétnych obchodných úlohách vrcholoví manažéri - tzv. HiPPO (obr. 2.1-9) - na základe skúseností a intuúcie, dnes, s prudkým nárastom objemu informácií, sa tento prístup stáva kontroverzným. Nahrádza ho automatizovaná analytika, ktorá pracuje s údajmi v reálnom čase.

"Tradično-manuálne" diskusie o podnikových procesoch na úrovni výkonných pracovníkov sa presunú k prevádzkovej analytike, ktorá si vyžaduje rýchle reakcie na obchodné otázky.

Éra, keď účtovníci, majstri a odhadcovia ručne vytvárali správy, súhrnné tabuľky a prehľadové tabuľky s údajmi o projekte počas dní a týždňov, je minulosťou. Dnes sa rýchlosť a včasnosť rozhodovania stáva klíčovým faktorom konkurenčnej výhody.



Obr. 4.2-1 Stavebníctvo potrebuje na výpočet a rozhodnutie niekoľko dní, na rozdiel od iných odvetví, kde sa to deje v priebehu hodín alebo minút.

Hlavným rozdielom medzi stavebníctvom a digitálne vyspelejšími odvetviami (obrázok 4.2-1) je nízka úroveň kvality a štandardizácie údajov. Zastarané prístupy k tvorbe, prenosu a spracovaniu informácií spomaľujú procesy a vytvárajú chaos. Nedostatok jednotných štandardov kvality údajov bráni implementácii komplexnej automatizácie.

Jednou z hlavných výziev zostáva nízka kvalita vstupných údajov, ako aj nedostatok formalizovaných postupov na ich prípravu a overovanie. Bez spoločlivých a konzistentných údajov nie je možná efektívna integrácia medzi systémami. To vedie k oneskoreniam, chybám a zvýšeným nákladom v každej fáze životného cyklu projektu.

V nasledujúcich častiach knihy sa podrobne venujeme tomu, ako môžete zlepšiť kvalitu údajov, štandardizovať procesy a skrátiť cestu od informácií ku kvalitným, overeným a konzistentným údajom.

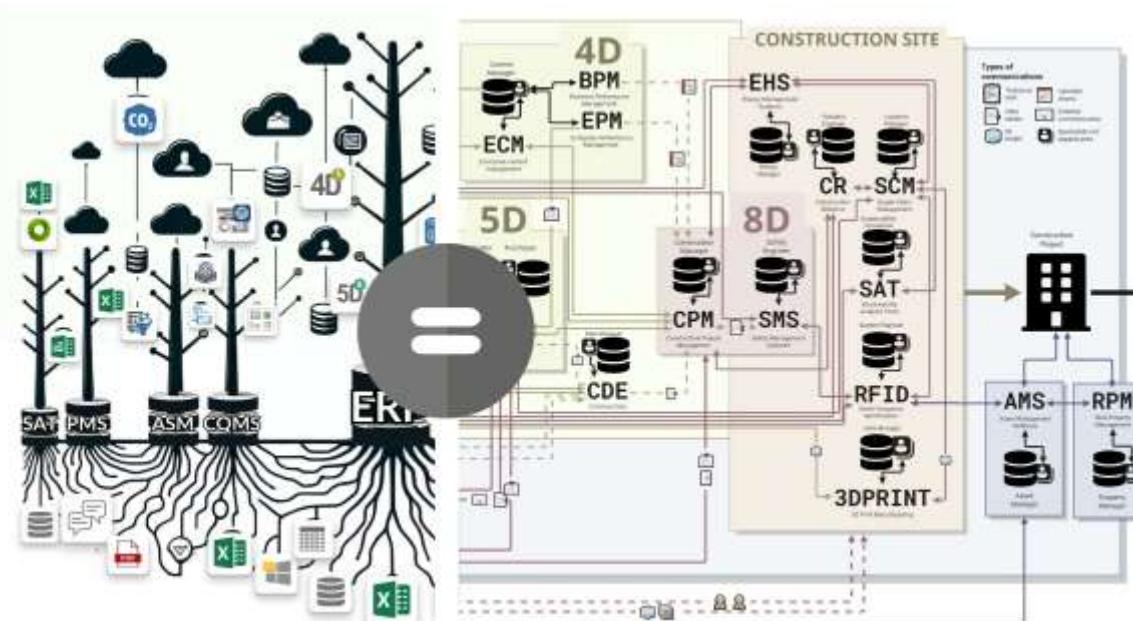
## Štandardizácia a integrácia údajov

Účinná správa údajov si vyžaduje jasného strategiu štandardizácie. Iba s jasnými požiadavkami na štruktúru a kvalitu údajov možno automatizovať validáciu údajov, obmedziť manuálne operácie a urýchliť informované rozhodovanie vo všetkých fázach projektu.

V každodennej praxi musí stavebná spoločnosť denne spracúvať stovky súborov: e-maily, dokumenty vo formáte PDF, návrhové súbory CAD, údaje zo senzorov IoT, ktoré je potrebné integrovať do obchodných procesov spoločnosti.

Les ekosystému databáz a nástrojov spoločnosti (obrázok 4.2-2) sa musí naučiť získavať živiny z prichádzajúcich multiformátových údajov, aby sa dosiahli výsledky, ktoré spoločnosť chce.

Na efektívne zvládnutie toku údajov nemusíte nevyhnutne zamestnávať armádu manažérov, ale musíte najprv vytvoriť prísne požiadavky a normy pre údaje a používať vhodné nástroje na ich automatické overovanie, harmonizáciu a spracovanie.



Obr. 4.2-2 Zabezpečenie zdravého ekosystému podniku si vyžaduje kvalitné a včasné zásobovanie jeho systémov.

S cieľom automatizovať proces validácie a harmonizácie údajov (pre následnú automatickú integráciu) by ste mali začať opisom minimálnych požiadaviek na údaje pre každý konkrétny systém. Tieto požiadavky definujú:

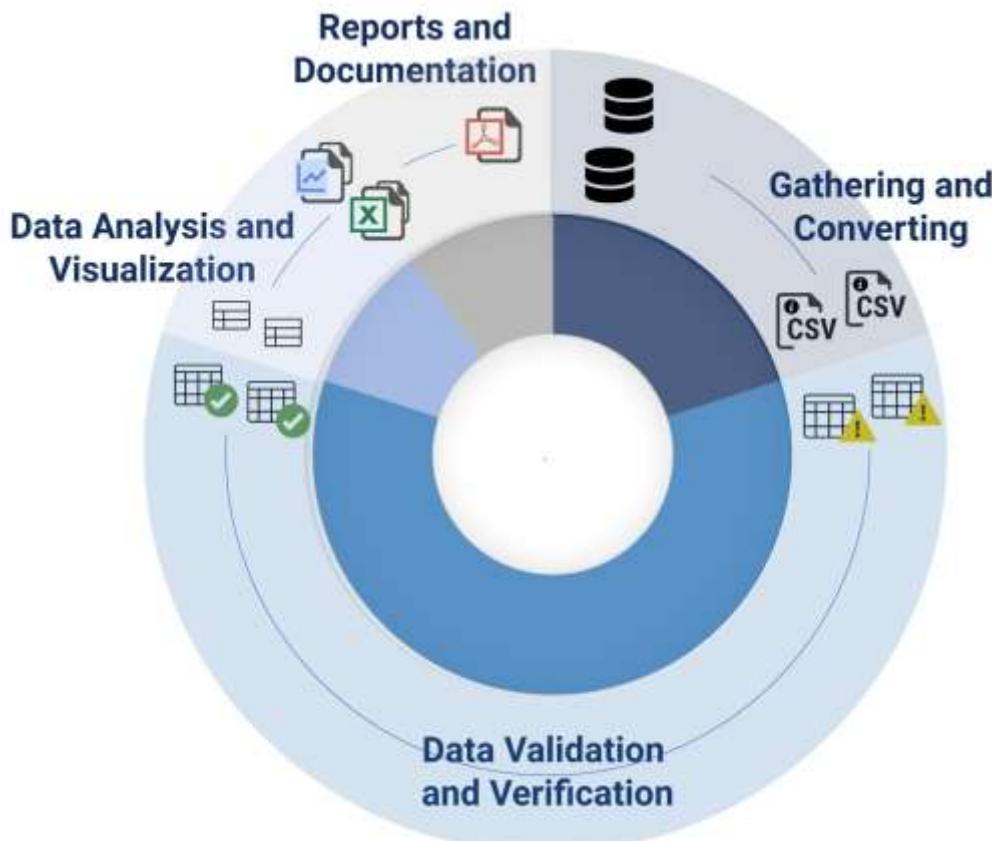
- Čo presne potrebujete získať?
- V akej forme (štruktúra, formát)?
- Aké atribúty sú povinné?
- Aké tolerancie presnosti a úplnosti sú prijateľné?

Požiadavky na údaje opisujú kritériá kvality, štruktúry a úplnosti prijatých a spracovaných informácií. Napríklad v prípade textov v PDF -dokumentoch je dôležité zabezpečiť presné formátovanie v súlade s priemyselnými normami (obr. 7.2-14 - obr. 7.2-16). Objekty v CAD -modeloch musia mať správne atribúty (rozmery, kódy, odkazy na klasifikátory) (obr. 7.3-9, obr. 7.3-10). A pri skenovaní zákaziek sú dôležité jasné dátumy a možnosť automatického extrahovania sumy a kľúčových pojmov (obr. 4.1-7 - obr. 4.1-10).

Formulovanie požiadaviek na údaje a automatická kontrola ich súladu je jedným z časovo najnáročnejších, ale rozhodujúcich krov. Je to časovo najnáročnejší krok v podnikových procesoch.

Ako bolo uvedené v tretej časti tejto knihy, 50% až 90% času profesionálov v oblasti business intelligence (BI) sa venuje skôr príprave údajov ako ich analýze (obrázok 3.2-5). Tento proces zahŕňa zber, overovanie, validáciu, harmonizáciu a štruktúrovanie údajov.

Podľa prieskumu z roku 2016 [95] dátoví vedci v najrôznejších širokospektrálnych oblastiach uviedli, že väčšinu svojho pracovného času (približne 80%) trávia tým, čo robia najmenej radi (obrázok 4.2-3): zhromažďovaním existujúcich súborov údajov a ich organizovaním (zjednocovaním, štruktúrovaním). Na tvorivé úlohy, ako je hľadanie zákonitostí a pravidelností, ktoré povedú k novým poznatkom a objavom, im tak zostáva menej ako 20% času.



Obr. 4.2-3 Overenie a zabezpečenie kvality údajov je najnákladnejším, časovo najnáročnejším a najzložitejším krokom pri príprave údajov na integráciu do iných systémov.

Úspešná správa údajov v stavebnej spoločnosti si vyžaduje komplexný prístup, ktorý zahrňa parametrizáciu úloh, formuláciu požiadaviek na kvalitu údajov a použitie vhodných nástrojov na ich automatizovanú validáciu.

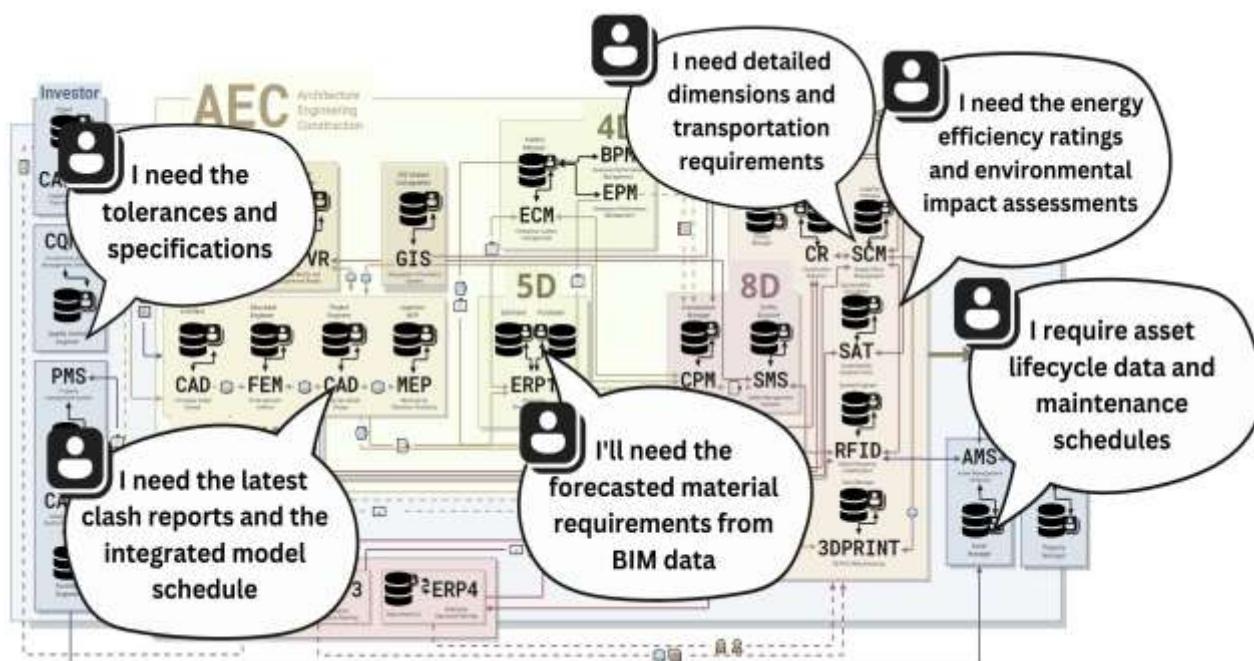
## Digitálna interoperabilita sa začína požiadavkami

S rastúcim počtom digitálnych systémov v spoločnostiach rastie aj potreba konzistentnosti údajov medzi nimi. Manažéri zodpovední za rôzne IT systémy často nestihajú sledovať rastúci objem informácií a rôznorodosť formátov. Za takýchto okolností sú nútení požiadať špecialistov o vytvorenie údajov vo forme vhodnej na použitie v iných aplikáciach a platformách.

To si zasa vyžaduje, aby sa inžinieri a zamestnanci zapojení do tvorby údajov prispôsobovali množstvu požiadaviek, často bez transparentnosti a jasného pochopenia toho, kde a ako sa budú údaje v budúcnosti používať. Nedostatok štandardizovaných prístupov k spracovaniu informácií vede k neefektívnosti a zvýšeným nákladom vo fáze overovania, ktorá je často manuálna z dôvodu zložitosti a neštandardizovanej povahy údajov.

Otázka štandardizácie údajov nie je len otázkou pohodlia alebo automatizácie. Ide o priame finančné straty. Podľa správy spoločnosti IBM z roku 2016 ročné straty spôsobené zlou kvalitou údajov v USA predstavujú 3,1 bilióna dolárov [96]. Okrem toho štúdie MIT a ďalších analytických konzultačných firiem ukazujú, že náklady na zlú kvalitu údajov môžu predstavovať až 15 - 25% príjmov spoločnosti [97].

Za týchto podmienok je veľmi dôležité mať jasne definované požiadavky na údaje a opisy toho, aké parametre, v akom formáte a s akou úrovňou podrobnosti by mali byť zahrnuté vo vytvorených objektoch. Bez formalizácie týchto požiadaviek nie je možné zaručiť kvalitu údajov a kompatibilitu medzi systémami a fázami projektu (obr. 4.2-4).

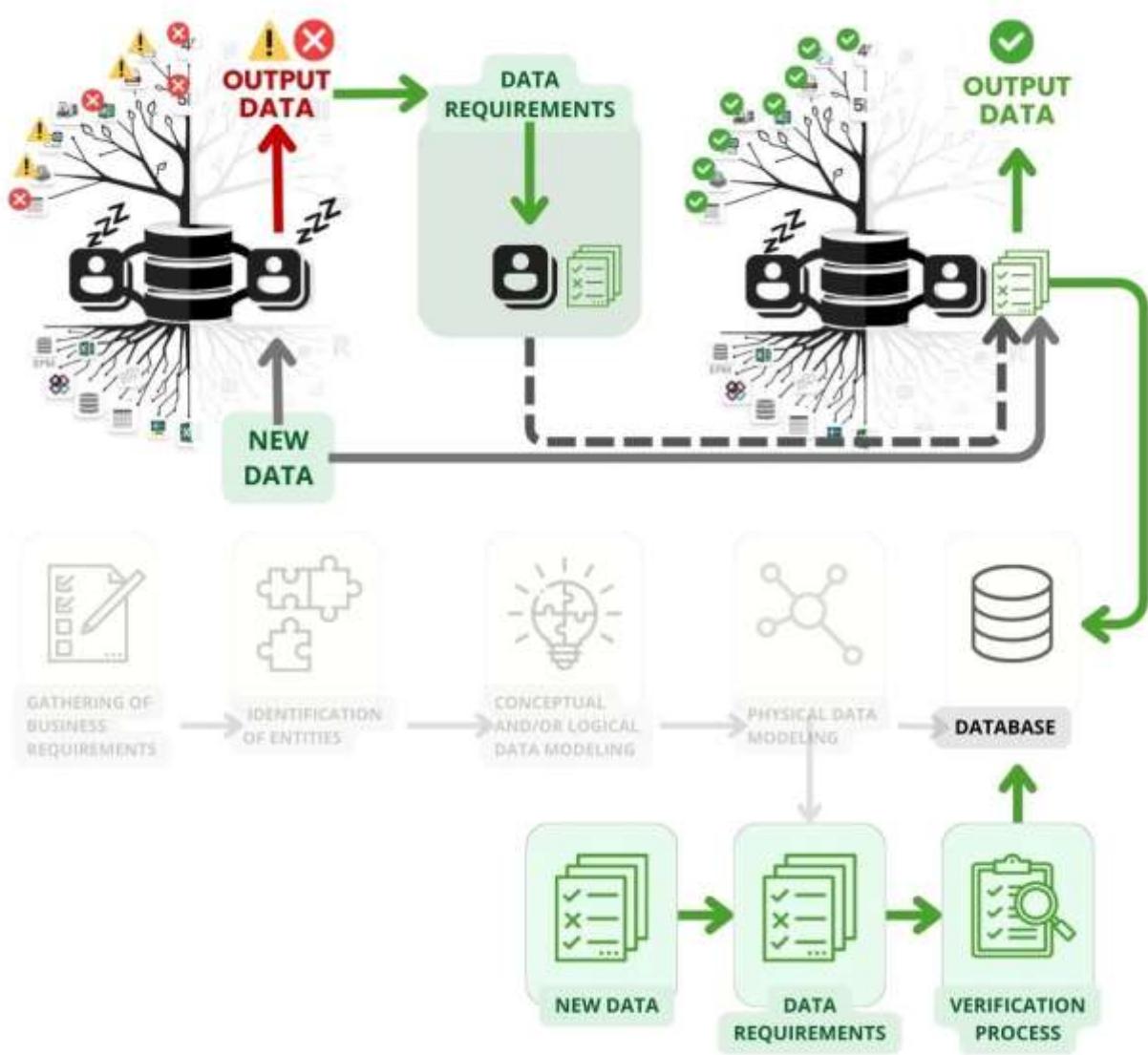


Obr. 4.2-4 Podnikanie je založené na interakcii rôznych rolí, z ktorých každá vyžaduje určité parametre a hodnoty, ktoré sú rozhodujúce pre naplnenie podnikateľských cieľov.

Aby ste mohli správne formulovať požiadavky na údaje, musíte pochopiť podnikové procesy na úrovni

údajov. Stavebné projekty sa líšia typom, rozsahom a počtom účastníkov a každý systém - či už ide o modelovanie (CAD (BIM)), plánovanie (ERP 4D), kalkuláciu nákladov (ERP 5D) alebo logistiku (SCM) - si vyžaduje vlastné jedinečné parametre vstupov (vstupné entity - prvky).

V závislosti od týchto potrieb musia podnikoví manažéri buď navrhnuť nové dátové štruktúry, aby splnili požiadavky, alebo prispôsobiť existujúce tabuľky a databázy. Kvalita vytvorených údajov bude priamo závisieť od toho, ako presne a správne budú požiadavky formulované (obr. 4.2-5).



Obrázok 4.2-5 Kvalita údajov závisí od kvality požiadaviek, ktoré sú vytvorené pre konkrétné prípady použitia údajov.

Kedže každý systém má svoje vlastné špecifické požiadavky na údaje, prvým krokom pri formulovaní všeobecných požiadaviek je kategorizácia všetkých prvkov zapojených do obchodných procesov. To znamená nutnosť rozdeliť objekty do tried a skupín tried zodpovedajúcich konkrétnym systémom alebo

aplikáčnym úlohám. Pre každú takúto skupinu sa vypracujú samostatné požiadavky na štruktúru, atribúty a kvalitu údajov.

V praxi však implementácia tohto prístupu čelí veľkej výzve: chýba spoločný jazyk na zoskupovanie údajov. Rozdielne klasifikácie, duplicitné identifikátory a nekompatibilné formáty vedú k tomu, že každá spoločnosť, každý softvér a dokonca každý projekt si vytvára vlastné, izolované dátové modely a triedy. Výsledkom je digitálna "babylonská veža", kde si prenos informácií medzi systémami vyžaduje viacnásobné konverzie na správne dátové modely a triedy, často vykonávané ručne. Túto bariéru možno prekonať len prechodom k univerzálnym klasifikátorom a štandardizovaným súborom požiadaviek.

## **Spoločný jazyk konštrukcie: úloha klasifikátorov v digitálnej transformácii**

V kontexte digitalizácie a automatizácie procesov kontroly a spracovania zohrávajú osobitnú úlohu klasifikačné systémy - akési "digitálne slovníky", ktoré zabezpečujú jednotnosť pri popise a parametrizácii objektov. Klasifikátory tvoria "spoločný jazyk", ktorý umožňuje zoskupovať údaje podľa významu a integrovať ich medzi rôznymi systémami, úrovňami riadenia a fázami životného cyklu projektu.

Najhmatateľnejší vplyv klasifikátorov je v ekonomike životného cyklu budovy, kde je najdôležitejším aspektom optimalizácia dlhodobých prevádzkových nákladov. Štúdie ukazujú, že prevádzkové náklady predstavujú až 80% celkových nákladov na vlastníctvo budovy, čo je trikrát viac ako počiatočné náklady na výstavbu (obr. 4.2-6) [98]. To znamená, že rozhodnutie o budúcich nákladoch sa vo veľkej mieri formuje vo fáze projektovania

Preto by sa požiadavky od prevádzkových inžinierov (CAFM, AMS, PMS, RPM) mali stať východiskom pre vytváranie požiadaviek na údaje vo fáze návrhu (obr. 1.2-4). Tieto systémy by sa nemali vnímať ako záverečná fáza projektu, ale ako neoddeliteľná súčasť celého digitálneho ekosystému projektu od koncepcie až po demontáž

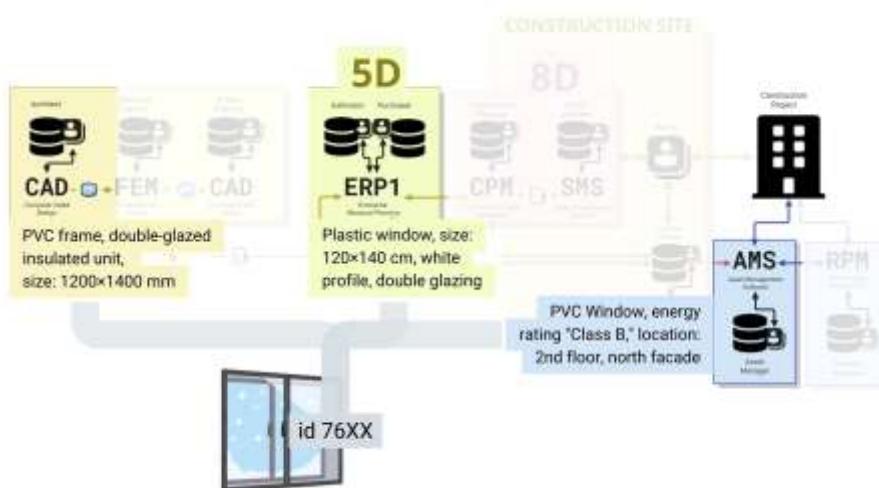
Moderný klasifikátor nie je len systém kódov na zoskupovanie. Je to mechanizmus na vzájomné porozumenie medzi architektmi, inžiniermi, odhadcami, logistikmi, údržbármi a IT systémami. Tak ako autopilot automobilu musí jednoznačne a presne rozpoznávať objekty na ceste, systémy digitálnej výstavby a ich používatelia musia ten istý prvok projektu jednoznačne interpretovať pre rôzne systémy prostredníctvom triedy prvkov.



Obr. 4.2-6 Náklady na prevádzku a údržbu trojnásobne prevyšujú náklady na výstavbu a predstavujú 60 až 80% celkových nákladov na životný cyklus budovy (podľa [99]).

Úroveň vývoja klasifikátora priamo súvisí s hĺbkou digitalizácie spoločnosti a jej digitálnou vyspelosťou. Organizácie s nízkou úrovňou digitálnej vyspelosti sa stretávajú s roztrieštenými údajmi, nekompatibilnými informačnými systémami a v dôsledku toho s nekompatibilnými a neefektívnymi klasifikátormi. V takýchto spoločnostiach môže mať ten istý prvak v rôznych systémoch často rôzne identifikátory zoskupenia, čo kriticky brzdí konečnú integráciu a znemožňuje automatizáciu procesov.

Napríklad to isté okno v projekte môže byť v CAD modeli, systéme odhadu a údržby (obr. 4.2-7) označené rôzne z dôvodu viacozmerného vnímania prvkov rôznymi účastníkmi procesu. Pre odhadcu sú v kategórii okna dôležité prvky, objem a náklady, pre servis údržby dostupnosť a udržiavateľnosť, pre architekta estetické a funkčné vlastnosti. Výsledkom je, že ten istý prvak môže vyžadovať rôzne parametre.



Obr. 4.2-7 Pri nekonzistentnej klasifikácii medzi systémami prvak stratí niektoré informácie o atribútoch v každom kroku svojho prechodu do iného systému.

Vzhľadom na náročnosť jednoznačného definovania klasifikácie stavebných prvkov odborníci z

rôznych oblastí často priraďujú tomu istému prvku nezlučiteľné triedy. To vede k strate jednotného pohľadu na objekt, čo si vyžaduje následné manuálne zásahy na zosúladenie rôznych klasifikačných systémov a vytvorenie konzistentnosti medzi typmi a triedami definovanými rôznymi odborníkmi.

V dôsledku tejto nejednotnosti sa často stáva, že operatívna dokumentácia, ktorú dostáva oddelenie verejného obstarávania (ERP) pri nákupe stavebnej položky od výrobcu, sa nedá správne prepojiť s klasifikáciou tejto položky na stavbe (PMIS, SCM). V dôsledku toho nie je pravdepodobné, že kritické informácie budú integrované do systémov riadenia infraštruktúry a majetku (CAFM, AMS), čo spôsobuje vážne problémy počas uvádzania do prevádzky, ako aj počas následnej údržby (AMS, RPM) alebo výmeny prvkmu.

V spoločnostiach s vysokou digitálnou vyspelosťou zohrávajú klasifikátory úlohu nervového systému, ktorý spája všetky informačné toky. Ten istý prvok dostáva jedinečný identifikátor, ktorý umožňuje jeho prenos medzi CAD, ERP, AMS a CAFM -systémami a ich klasifikátormi bez skreslenia alebo straty.

Ak chcete vytvoriť účinné klasifikátory, musíte pochopiť, ako sa údaje používajú. Ten istý inžinier môže v rôznych projektoch pomenovať a klasifikovať prvok odlišne. Stabilný klasifikačný systém sa dá vytvoriť len zhromažďovaním štatistik o používaní v priebehu rokov. Pomáha pri tom strojové učenie: algoritmy analyzujú tisíce projektov (obr. 9.1-10) a pomocou strojového učenia identifikujú pravdepodobné triedy a parametre (obr. 10.1-6). Automatická klasifikácia je obzvlášť cenná v prostredí, kde manuálna klasifikácia nie je možná vzhľadom na objem údajov. Automatické klasifikačné systémy budú schopné rozlišovať základné kategórie na základe minimálne vyplnených parametrov položiek (podrobnejšie v deviatej a desiatej časti knihy).

Vyvinuté klasifikačné systémy sa stávajú katalyzátorom ďalšej digitalizácie a vytvárajú základ pre:

- Automatizovaný odhad nákladov a časového harmonogramu projektu.
- Prediktívna analýza potenciálnych rizík a konfliktov
- Optimalizácia procesov obstarávania a logistických reťazcov
- Vytváranie digitálnych dvojčiat budov a štruktúr
- Integrácia so systémami inteligentného mesta a internetu vecí

Čas na transformáciu je obmedzený - s rozvojom technológií strojového učenia a počítačového videnia sa problém automatickej klasifikácie, ktorý bol desaťročia neriešiteľný, v najbližších rokoch vyrieši a stavebné a projektové spoločnosti, ktoré sa včas neprispôsobia, riskujú, že sa zopakuje osud vozových parkov taxíkov vytlačených digitálnymi platformami.

Automatizácia výpočtov nákladov a harmonogramu, ako aj veľké dátá a strojové učenie budú podrobnejšie rozobrané v piatej a deviatej časti knihy. Riziku zopakovania osudu vozových parkov taxislužieb a uberizácie stavebníctva sa podrobne venuje desiatka časť knihy.

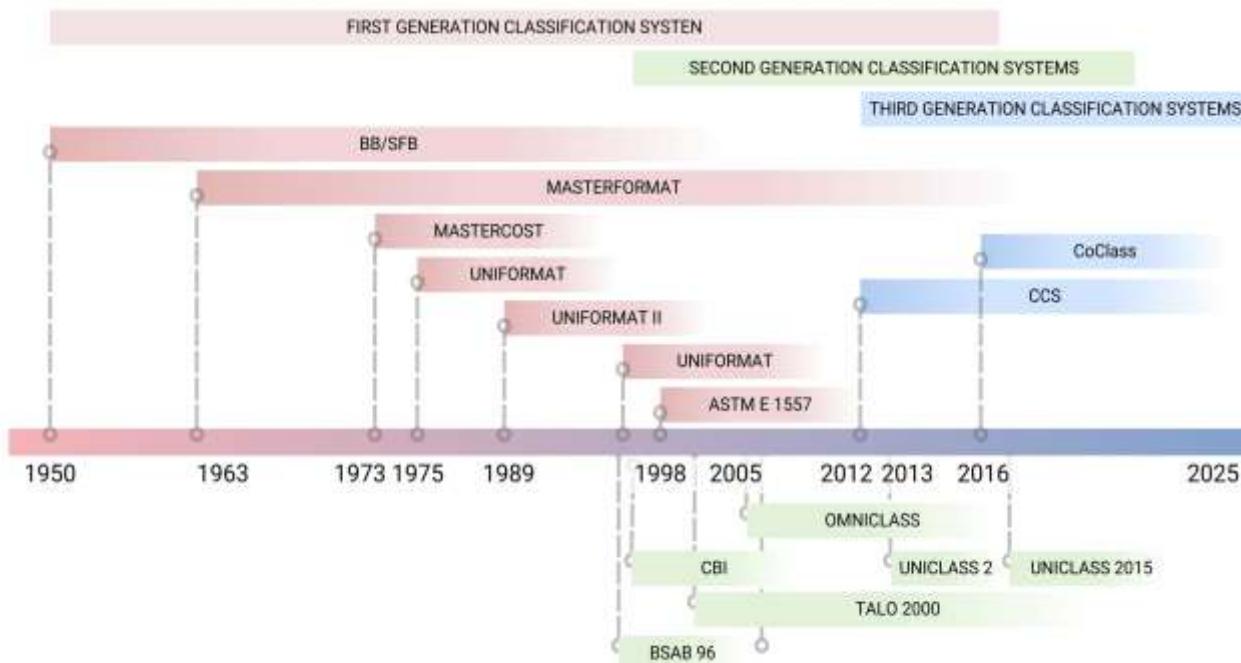
Pri pochopení kľúčovej úlohy klasifikátorov v digitálnej transformácii stavebného priemyslu je potrebné

venovať sa historii ich vývoja. Práve historický kontext nám umožňuje uvedomiť si, ako sa prístupy ku klasifikátorom vyvíjali a aké trendy definujú ich súčasný stav.

## Masterformat, OmniClass, Uniclass a CoClass: vývoj klasifikačných systémov

Historicky sa klasifikátory stavebných prvkov a práv vyvíjali v troch generáciách, pričom každá z nich odrážala úroveň dostupnej technológie a aktuálne potreby odvetvia v určitom časovom období (obrázok 4.2-8):

- **Prvá generácia** (začiatok 50. rokov až koniec 80. rokov) - papierové adresáre, hierarchické klasifikátory používané lokálne (napr. Masterformat, SfB).
- **Druhá generácia** (koniec 90. rokov až polovica roka 2010) sú tabuľky a štruktúrované databázy implementované v Exceli a Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Tretia generácia** (od roku 2010 po súčasnosť) - digitálne služby a API -rozhrania, integrácia s CAD (BIM), automatizácia (Uniclass 2015, CoClass).



Obr. 4.2-8 Tri generácie klasifikátorov pre stavebníctvo.

V posledných desaťročiach došlo k zníženiu hierarchickej zložitosti (obr. 4.2-9) klasifikátorov: zatiaľ čo prvé systémy, ako napríklad OmniClass, používali až 7 úrovňí vnorenia na opis 6887 tried, moderné riešenia, ako napríklad CoClass, sa obmedzujú na 3 úrovne so 750 triedami. To uľahčuje prácu s údajmi pri zachovaní potrebnej granularity. Uniclass 2015, ktorý sa často používa ako štandard v Spojenom kráľovstve, spája 7210 tried len v 4 úrovniach, čo je vhodné pre projekty CAD a verejné obstarávanie.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Obr. 4.2-9 S každou novou generáciou klasifikátorov sa zložitosť kategorizácie dramaticky znižuje.

V systémoch odhadu stavebných prác v rôznych krajinách sa v dôsledku rozdielov v klasifikácii môže aj typický prvak, ako je betónová základová stena, opísať veľmi odlišným spôsobom (obrázok 4.2-10). Tieto rozdiely odrážajú národné stavebné postupy, používané meracie systémy, prístupy ku klasifikácii materiálov a regulačné a technické požiadavky platné v jednotlivých krajinách.



Obr. 4.2-10 Ten istý prvak sa v projektoch v rôznych krajinách používa prostredníctvom rôznych opisov a klasifikácií.

Rôzne klasifikácie tých istých prvkov komplikujú medzinárodnú spoluprácu a spôsobujú, že porovnanie nákladov a rozsahu prác v rámci medzinárodných projektov je časovo náročné a niekedy takmer nemožné. V súčasnosti neexistuje jeden univerzálny klasifikátor na celosvetovej úrovni - každá krajina alebo región si vytvára vlastné systémy založené na miestnych normách, jazyku a obchodnej kultúre:

- **CCS** (Dánsko): Systém klasifikácie nákladov - systém klasifikácie nákladov počas celého životného cyklu zariadenia (projekt, výstavba, prevádzka). Dôraz sa kladie na logiku prevádzky a údržby, ale zahŕňa aj rozpočet a riadenie zdrojov.
- **NS 3451** (Nórsko): kategorizuje zariadenia podľa funkcie, konštrukčných prvkov a fáz životného cyklu. Používa sa na riadenie projektov, odhad nákladov a dlhodobé plánovanie.
- **MasterFormat** (USA): systém na štruktúrovanie stavebných špecifikácií do sekcií (napr. betón, elektroinštalačia, dokončovacie práce). Zameriava sa skôr na disciplíny a typy prác než na

funkčné prvky (na rozdiel od UniFormatu).

- **Uniclass 2** (UK): jeden z najpodrobnejších klasifikátorov, ktorý sa používa vo verejnom obstarávaní a v projektoch BIM. Zjednocuje údaje o objektoch, prácach, materiáloch a priestoroch do jedného systému.
- **OmniClass**: medzinárodný štandard (vyvinutý CSI v USA) na správu informácií o objektoch z knižníc komponentov do elektronických špecifikácií. Vhodný na dlhodobé ukladanie údajov, kompatibilný s CAD (BIM) a inými digitálnymi nástrojmi.
- **COBie**: Construction-Operation Building information exchange je medzinárodný štandard na výmenu údajov medzi fázami projektovania, výstavby a prevádzky. Zahrnutý do normy BS 1192-4:2014 ako súčasť koncepcie "BIM -model pripravený na použitie". Zameriava sa na prenos informácií (napr. špecifikácie zariadení, záruky, kontakty na dodávateľov).

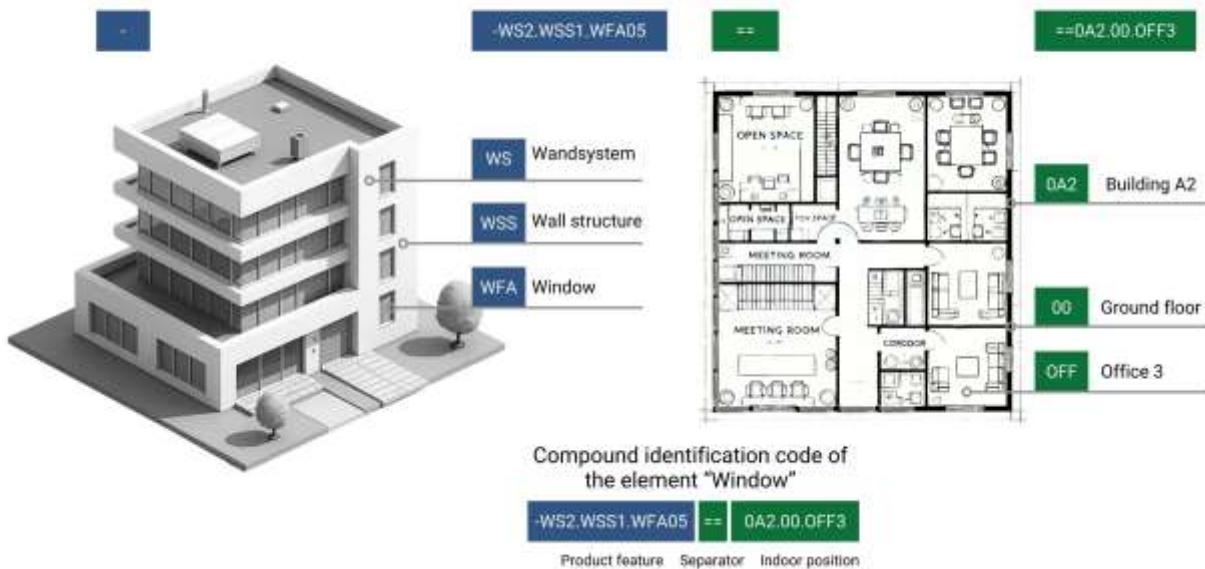
Globalizácia stavebníctva pravdepodobne povedie k postupnému zjednoteniu systémov klasifikácie stavebných prvkov, čo výrazne zníži závislosť od miestnych národných noriem. Tento proces sa môže vyvíjať podľa vzoru vývoja internetovej komunikácie, kde univerzálne protokoly na prenos údajov nakoniec nahradili rôznorodé miestne formáty a zabezpečili globálnu interoperabilitu systémov.

Alternatívou cestou vývoja môže byť priamy prechod na automatické klasifikačné systémy založené na technológiách strojového učenia. Tieto technológie, ktoré sa dnes vyvíjajú najmä v oblasti autonómnej dopravy, majú značný potenciál pre aplikáciu na veľké súbory údajov o dizajne CAD (obr. 10.1-6).

V súčasnosti sa situácia neobmedzuje len na národné zoskupenia klasifikátorov. Z dôvodu mnohých osobitostí, ktoré sa nezohľadňujú na národnej úrovni, musí každá spoločnosť zjednotiť a štandardizovať kategórie prvkov a zdrojov, s ktorými pracuje.

Tento proces sa spravidla začína v malom - s lokálnymi tabuľkami objektov alebo internými systémami označovania. Strategickým cieľom je však prejsť na spoločný jazyk na opis všetkých prvkov, ktorý by bol zrozumiteľný nielen v rámci podniku, ale aj mimo neho - v ideálnom prípade harmonizovaný s medzinárodnými alebo odvetvovými klasifikátormi (obr. 4.2-8). Tento prístup uľahčuje integráciu s externými partnermi, digitálnymi systémami a podporuje tvorbu jednotných end-to-end procesov v rámci životného cyklu objektov.

Pred prechodom na automatizáciu a škálovateľné IT systémy je potrebné buď použiť klasifikátory na národnej úrovni, alebo vytvoriť vlastnú, logickú a jednoznačnú štruktúru identifikácie prvkov. Každý objekt - či už je to okno (obrázok 4.2-11), dvere alebo technický systém - musí byť opísaný tak, aby ho bolo možné jednoznačne rozpoznať v akomkoľvek digitálnom systéme spoločnosti. To je rozhodujúce pri prechode od plochých výkresov k digitálnym modelom, ktoré pokrývajú fázu projektovania aj prevádzky budovy.



Obr. 4.2-11 Príklad zloženého identifikátora okenného prvku budovy na základe klasifikácie a polohy v budove.

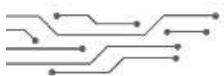
Jedným z príkladov interných klasifikátorov môže byť vytvorenie zloženého identifikačného kódu (obrázok 4.2-11). Takýto kód kombinuje niekoľko úrovni informácií: funkčný účel prvku (napr. "okno v stene"), jeho typ a presný priestorový odkaz - budova A2, poschodie 0, miestnosť 3. Takáto viacúrovňová štruktúra umožňuje vytvoriť jednotný systém navigácie v digitálnych modeloch a dokumentácii, najmä vo fázach overovania a transformácie údajov, kde sa vyžaduje jednoznačné zoskupenie prvkov. Jednoznačné rozpoznávanie prvkov zabezpečuje konzistentnosť medzi oddeleniami a znižuje riziko duplicity, chýb a straty informácií.

Dobre vytvorený klasifikátor nie je len technický dokument, ale základ digitálneho ekosystému spoločnosti:

- zabezpečuje kompatibilitu údajov medzi systémami;
- znižuje náklady na vyhľadávanie a spracovanie informácií;
- zvyšuje transparentnosť a ovládateľnosť;
- vytvára základ pre škálovanie a automatizáciu.

Štandardizovaný opis objektov prostredníctvom národných klasifikátorov alebo vlastných zložených identifikačných kódov sa stáva základom pre konzistentné údaje, spoľahlivú výmenu informácií a následnú implementáciu inteligentných služieb - od automatizovaného obstarávania až po digitálne dvojčičky.

Po dokončení fázy štruktúrovania viacformátových údajov a výbere klasifikátora, ktorý sa použije na rozpoznávanie a zoskupovanie prvkov, je ďalším krokom správne modelovanie údajov. Tento proces zahŕňa identifikáciu kľúčových parametrov, vytvorenie logickej štruktúry údajov a opis vzťahov medzi



prvkami.

## KAPITOLA 4.3.

# MODELOVANIE ÚDAJOV A CENTRUM EXCELENTNOSTI

### Modelovanie údajov: koncepčný, logický a fyzický model

Efektívna správa údajov (ktoré sme predtým štruktúrovali a kategorizovali) nie je možná bez premyslenej štruktúry ukladania a spracovania. Na zabezpečenie prístupu k informáciám a ich konzistentnosti vo fáze ukladania a spracovania využívajú spoločnosti dátové modelovanie, metodiku, ktorá im umožňuje navrhnuť tabuľky, databázy a väzby medzi nimi v súlade s obchodnými požiadavkami.

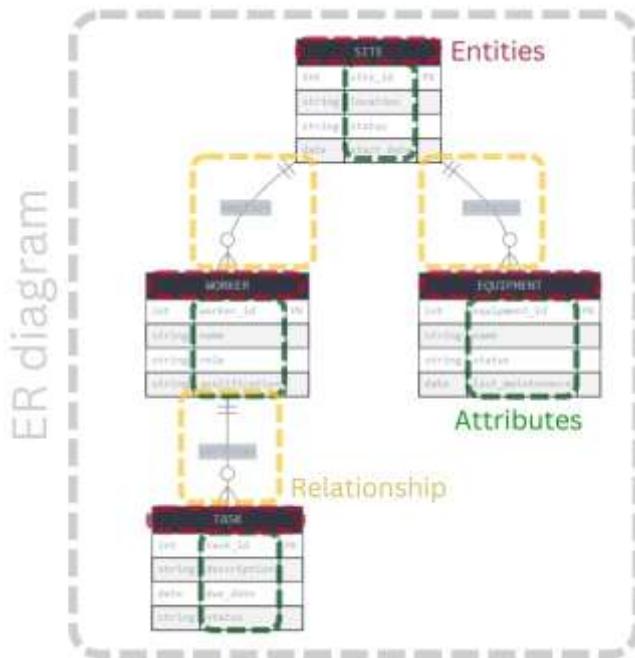
Modelovanie údajov je základom, na ktorom je postavený každý digitálny ekosystém. Bez opisu systémov, požiadaviek a dátového modelovania inžinieri a odborníci vytvárajúci údaje nevedia alebo nerozumejú, kde sa budú nimi vytvorené údaje používať.

Podobne ako pri stavbe budovy, kde nemôžete začať klásť tehly bez plánu, aj pri vytváraní systému dátového skladu je potrebné jasne pochopiť, aké údaje sa budú používať, ako budú prepojené a kto s nimi bude pracovať. Bez opisu procesov a požiadaviek strácajú inžinieri a odborníci vytvárajúci údaje prehľad o tom, kde a ako sa budú údaje v budúcnosti používať.

Dátový model slúži ako most medzi biznisom a IT. Umožňuje formalizovať požiadavky, štruktúrovať informácie a uľahčiť komunikáciu medzi zainteresovanými stranami. V tomto zmysle sa dátové modelovanie podobá práci architekta, ktorý podľa plánu zákazníka vypracuje plán budovy a potom ho odovzdá staveľom - správcom a vývojárom databázy - na realizáciu (vytvorenie databázy).

Každá stavebná firma tak musí okrem štruktúrovania a kategorizácie prvkov a zdrojov (obr. 4.2-11) ovládať aj umenie "budovania" databáz (tabuľiek) a naučiť sa vytvárať medzi nimi väzby, akoby spájať tehličky do spoľahlivej a pevnej steny znalostí z firemných údajov. Medzi kľúčové pojmy v modelovaní údajov (obrázok 4.3-1) patria:

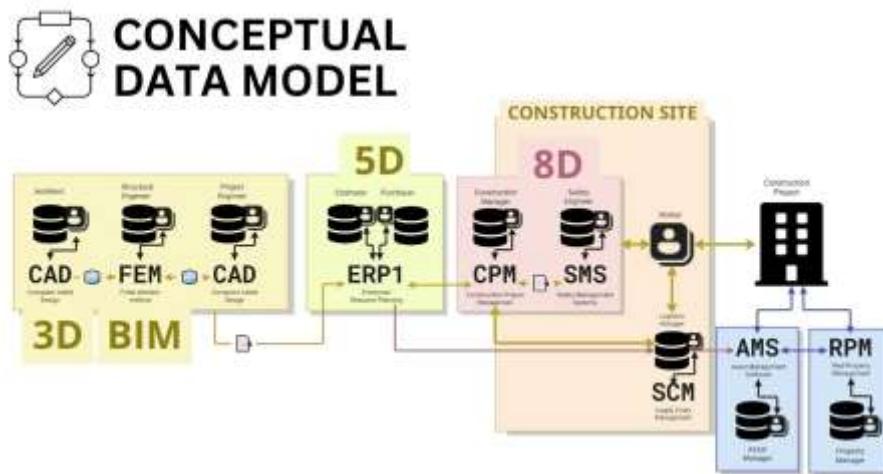
- **Entity** sú objekty, o ktorých sa musia zhromažďovať údaje. V počiatočnej fáze návrhu môže byť entitou jeden prvak (napr. "dvere") a v modeli odhadu to môže byť skupina prvkov zaradených do kategórií (napr. "vnútorné dvere").
- **Atribúty** sú charakteristiky entít, ktoré opisujú dôležité detaily: rozmer, vlastnosti, náklady na montáž, logistiku a ďalšie parametre.
- **Vzťahy (prepojenia)** - ukazujú, ako na seba entity vzájomne pôsobia. Môžu byť jedného z typov: "jeden k jednému", "mnoho k jednému", "mnoho k mnohým".
- **ER diagramy** (Entity-Relationship diagramy) sú vizuálne diagramy, ktoré zobrazujú entity, atribúty a vzťahy medzi nimi. ER-diagramy môžu byť konceptuálne, logické a fyzické - každý z nich odráža inú úroveň podrobnosti.



Obrázok 4.3-1 ER diagram konceptuálnej štruktúry databázy s entitami, atribútmi a vzťahmi.

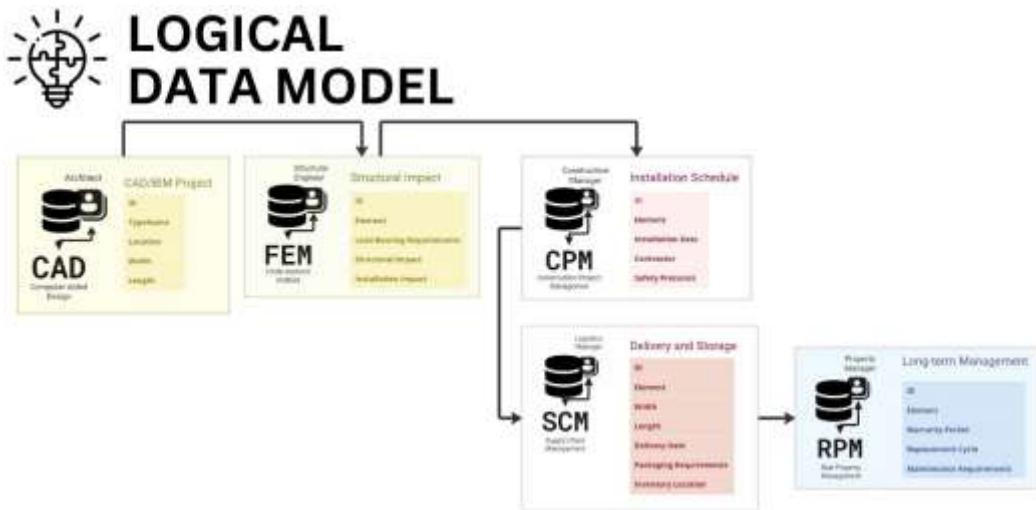
Proces navrhovania údajov a definovania vzťahov medzi nimi sa tradične delí na tri hlavné modely. Každý z nich plní určité funkcie, pričom sa líšia úrovňou podrobnosti a stupňom abstrakcie pri reprezentácii štruktúry údajov:

- **Konceptuálny dátový model:** tento model popisuje hlavné entity a ich vzťahy bez toho, aby zachádzal do detailov atribútov. Zvyčajne sa používa v počiatočných fázach plánovania. V tejto fáze môžeme z databáz a systémov načrtnúť náčrt, ktorý zobrazí vzťahy medzi rôznymi oddeleniami a špecialistami.



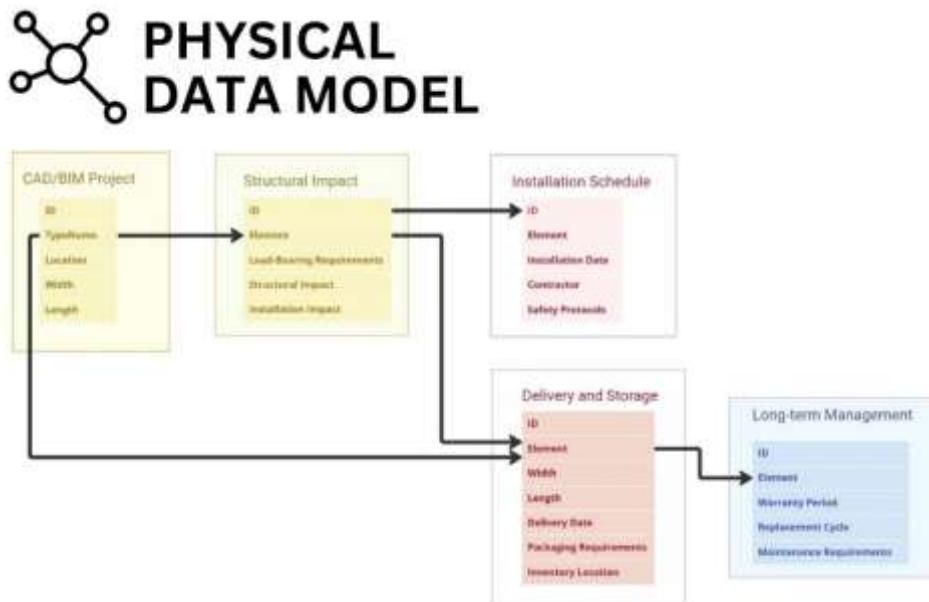
Obr. 4.3-2 Konceptuálny diagram popisuje obsah systému: vysokoúrovňové znázornenie vzťahov bez technických detailov.

- **Logický dátový model:** Na základe konceptuálneho modelu obsahuje logický dátový model podrobný popis entít, atribútov, klúčov a vzťahov, mapovanie obchodných informácií a pravidiel.



Obrázok 4.3-3 Logický dátový model podrobne opisuje dátové typy, vztahy a kľúče, ale bez systémovej implementácie.

- Fyzický dátový model:** Tento model opisuje štruktúry potrebné na implementáciu databázy vrátane tabuľiek, stĺpcov a vztáhov. Zameriava sa na výkonnosť databázy, stratégie indexovania a fyzické ukladanie s cieľom optimalizovať fyzické nasadenie databáz.

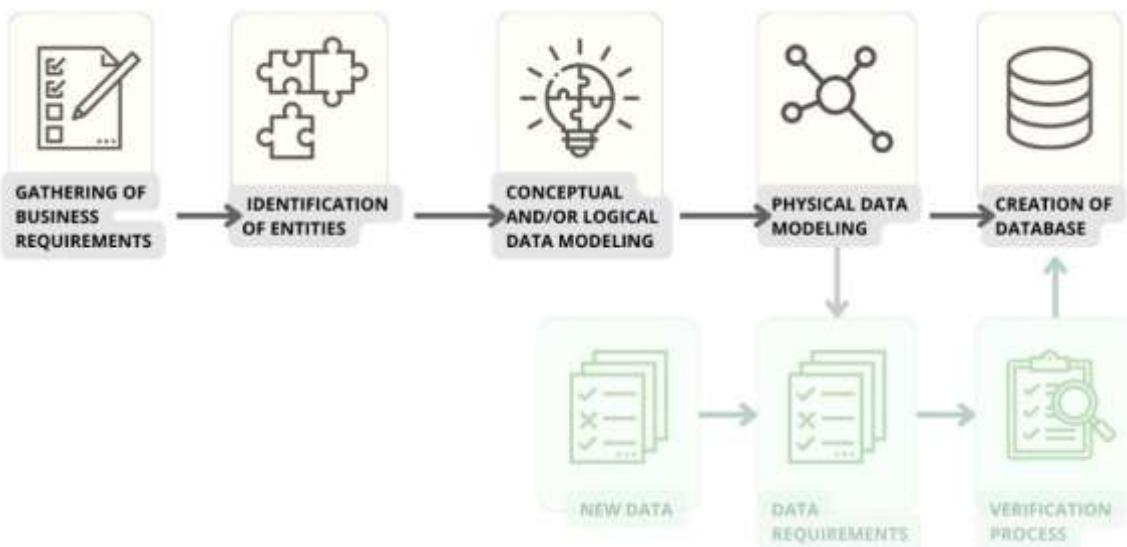


Obrázok 4.3-4 Fyzický dátový model definuje, ako bude systém implementovaný, vrátane tabuľiek a konkrétnych detailov databázy.

Pri návrhu databáz a návrhu tabuľkových vztáhov zohráva pochopenie úrovní abstrakcie kľúčovú úlohu pri vytváraní efektívnej architektúry systému.

Efektívna metodika modelovania údajov umožňuje spojiť obchodné ciele s technickou implementáciou, čím sa celý procesný reťazec stáva transparentnejším a lepšie zvládnuteľným. Modelovanie údajov nie je jednorazová úloha, ale proces zahŕňajúci postupné kroky (obrázok 4.3-5):

- **Zhromažďovanie obchodných požiadaviek:** definujú sa klúčové úlohy, ciele a informačné toky. V tejto fáze dochádza k aktívnej interakcii s odborníkmi a používateľmi.
- **Identifikácia entít:** zvýrazňujú sa hlavné objekty, kategórie a typy údajov, ktoré je dôležité zohľadniť v budúcom systéme.
- **Vývoj koncepcného a logického modelu:** najprv sa zachytia klúčové entity a ich vzťahy, potom atribúty, pravidlá a podrobnejšia štruktúra.
- **Fyzické modelovanie:** navrhuje sa technická implementácia modelu: tabuľky, polia, vzťahy, obmedzenia, indexy.
- **Vytvorenie databázy:** posledným krokom je implementácia fyzického modelu vo vybranom DBMS, testovanie a príprava na prevádzku.



Obr. 4.3-5 Tvorba databáz a systémov správy údajov pre podnikové procesy začína generovaním požiadaviek a modelovaním údajov.

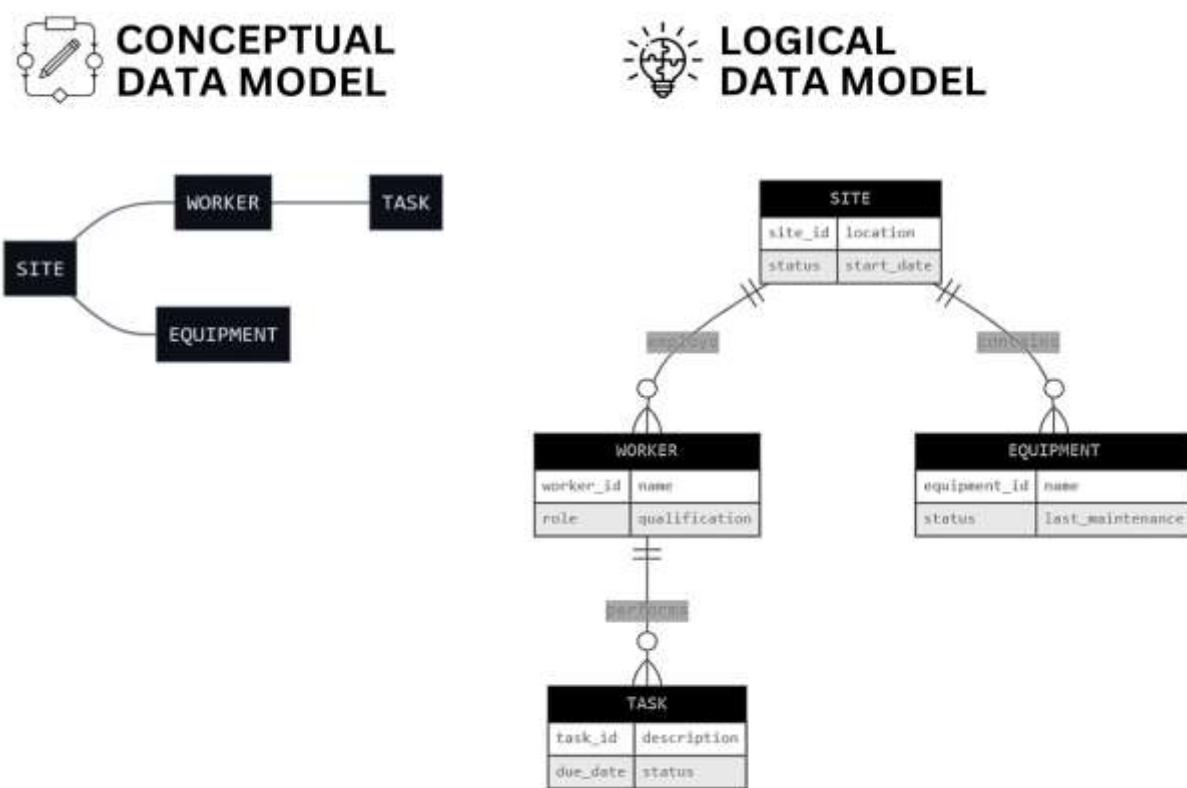
Správne navrhnuté procesy modelovania údajov umožňujú transparentné informačné toky, čo je dôležité najmä pri komplexných projektoch, ako je riadenie stavebných projektov alebo staveniska. Pozrime sa, ako môže prechod od koncepcného modelu k logickému modelu a následne k fyzickému modelu pomôcť zefektívniť procesy.

## Praktické modelovanie údajov v kontexte stavebníctva

Vezmieme si ako príklad modelovania údajov úlohu riadenia staveniska a prevedieme požiadavky majstra

do štruktúrovaného logického modelu. Na základe základných potrieb riadenia staveniska definujeme kľúčové entity pre: stavenisko (SITE), pracovníkov (WORKER), zariadenie (EQUIPMENT), úlohy (TASK) a použitie zariadenia (EQUIPMENT\_USAGE). Každá entita obsahuje súbor atribútov, ktoré odrážajú dôležité charakteristiky. Napríklad v prípade ÚLOHY to môže byť opis úlohy, dátum splatnosti, stav, priorita; v prípade PRACOVNÍKA to môže byť meno, jeho úloha na pracovisku, súčasné zamestnanie atď.

Logický model vytvára vzťahy medzi týmito entitami a ukazuje, ako sa navzájom ovplyvňujú v reálnych pracovných procesoch (obrázok 4.3-6). Napríklad vzťah medzi pracoviskom a pracovníkmi naznačuje, že na jednom pracovisku môže pracovať veľa pracovníkov, zatiaľ čo vzťah medzi pracovníkmi a úlohami vyjadruje, že jeden pracovník môže vykonávať viacero úloh.



Obr. 4.3-6 Konceptuálny a logický dátový model vytvorený na základe požiadaviek majstra na popis procesov na stavenisku.

Pri prechode na fyzický model sa pridávajú technické implementačné detaily: špecifické typy údajov (VARCHAR, INT, DATE), primárne a cudzie kľúče pre vzťahy medzi tabuľkami a indexy na optimalizáciu výkonu databázy (obrázok 4.3-7).

Napríklad pre stavy by sa mali definovať špecifické typy s možnými hodnotami a mali by sa pridať indexy na kľúčové polia, ako je status a worker\_id, aby sa zlepšil výkon vyhľadávania. Tým sa logický opis systému zmení na konkrétny plán implementácie databázy, ktorý je pripravený na vytvorenie a implementáciu.

# PHYSICAL DATA MODEL



Obrázok 4.3-7 Fyzický dátový model opisuje entity staveniska prostredníctvom minimálnych požadovaných parametrov.

Fyzický model sa často líši od logického modelu. V priemere je rozdelenie času modelovania nasledovné: približne 50% sa venuje konceptuálnemu modelu (zber požiadaviek, diskusia o procesoch, identifikácia entít), 10% logickému modelu (špecifikácia atribútov a vzťahov) a 40% fyzickému modelu (implementácia, testovanie, prispôsobenie DBMS).

Táto rovnováha je spôsobená tým, že konceptuálna fáza vytvára základ dátovej štruktúry, zatiaľ čo logický model len špecifikuje vzťahy a atribúty. Fyzický model si vyžaduje najviac zdrojov, pretože práve v tejto fáze sa údaje implementujú do konkrétnych platform a nástrojov.

## Vytvorenie databázy pomocou LLM

Po vytvorení dátového modelu a popisu entít prostredníctvom parametrov sme pripravení na vytvorenie databáz - úložísk, kde budeme ukladať informácie prichádzajúce po fáze štruktúrovania na konkréne procesy.

Pokúsime sa vytvoriť príklad jednoduchej, ale funkčnej databázy s minimálnym množstvom kódu pomocou SQLite na príklade programovacieho jazyka Python. Relačné databázy sme podrobne rozoberali v kapitole "Štruktúrované relačné databázy a dotazovací jazyk SQL".

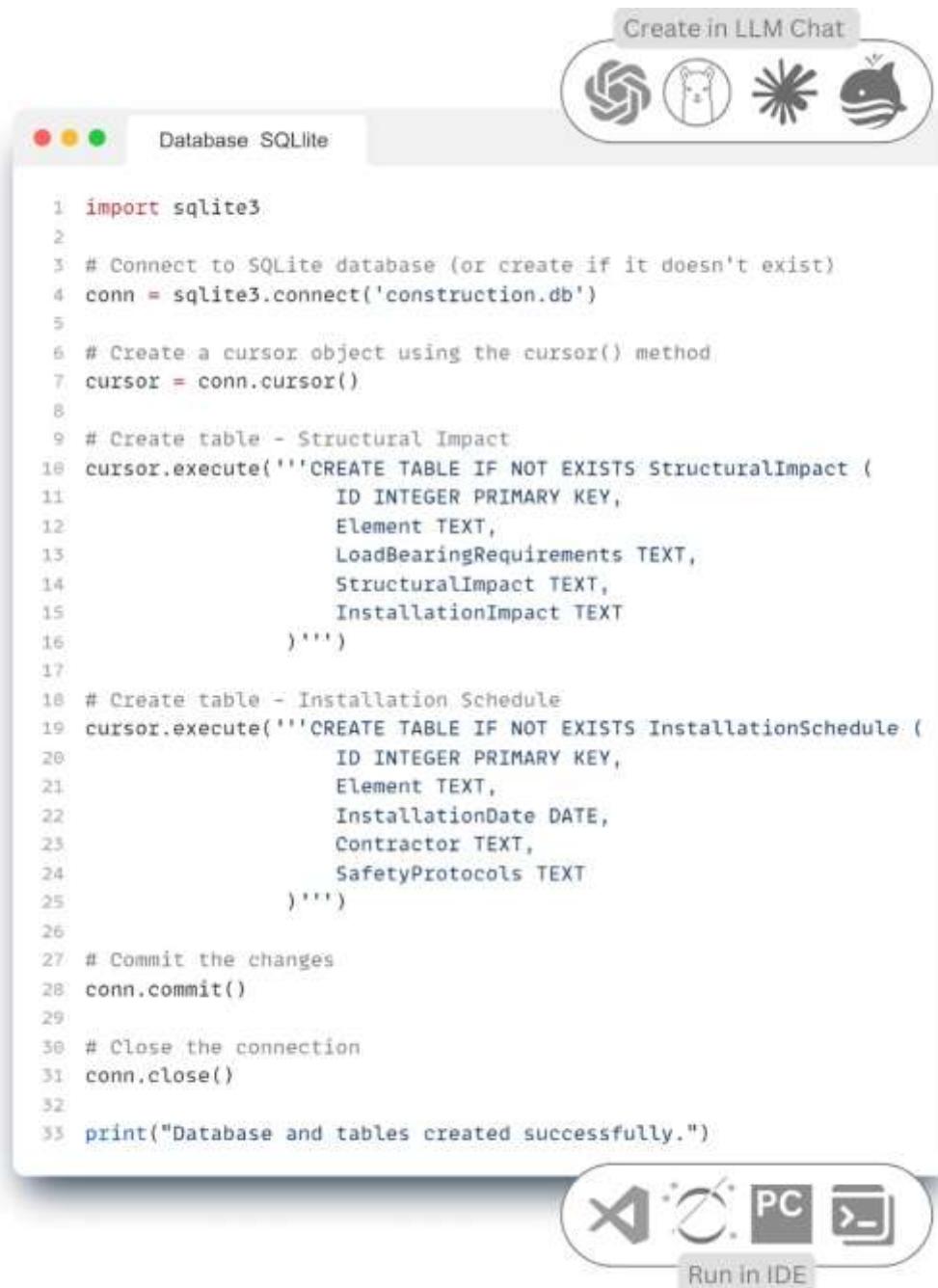
SQLite je ľahká, open source, vložiteľná, multiplatformná databáza, ktorá podporuje základné funkcie SQL. SQLite sa široko používa v CAD (BIM), ERP, EPM a je prítomný takmer v každom smartfóne.

Požiadajme model LLM, aby pre nás vytvoril malú databázu SQLite na základe vyššie vytvoreného fyzického modelu (obrázok 4.3-4) pre systémy FEM (vplyv konštrukcie) a CPM (plán montáže).

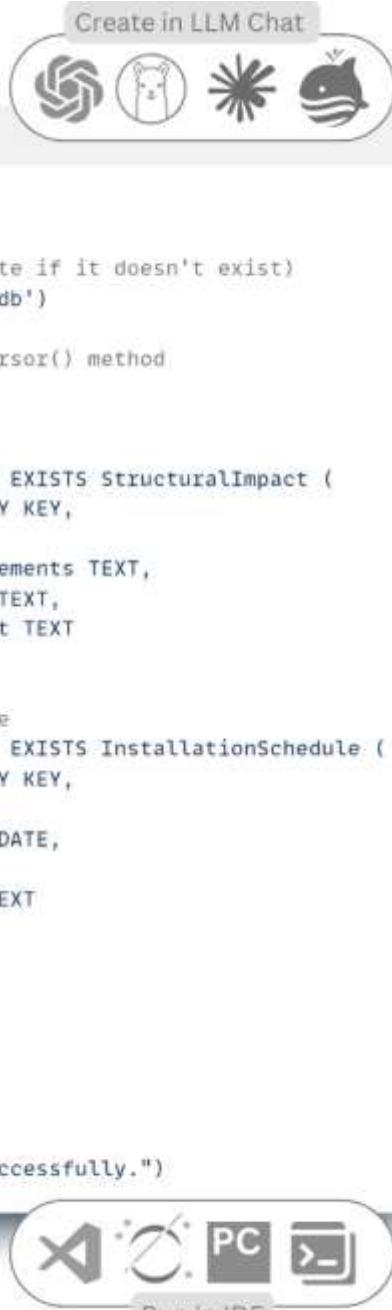
- ❷ Pošlite textovú požiadavku na chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akýkoľvek iný):

Vytvorte databázy SQLite pre systémy "Konštruktívny vplyv" a "Plán inštalácie" z obrázku v prílohe (vložte schému do chatu ako obrázok (obr. 4.3-4) ↴

💡 Odpoveď LLM:



```
1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11                 ID INTEGER PRIMARY KEY,
12                 Element TEXT,
13                 LoadbearingRequirements TEXT,
14                 StructuralImpact TEXT,
15                 InstallationImpact TEXT
16             )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20                 ID INTEGER PRIMARY KEY,
21                 Element TEXT,
22                 InstallationDate DATE,
23                 Contractor TEXT,
24                 SafetyProtocols TEXT
25             )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")
```



Obrázok 4.3-8 Vygenerovaný kód vytvorí z nahraného obrázka databázy a tabuľky, do ktorých teraz môžeme načítať údaje.

Tento databázový kód možno spustiť v prostredí Python s podporou SQLite, ktorá je zvyčajne súčasťou štandardného prostredia Python. Možno ho spustiť na ľubovoľnom lokálnom počítači s jazykom Python schopnou spúštať skripty a vytvárať súbory. Po spustení a vykonaní tohto kódu sa na pevnom disku vytvorí súbor - databáza SQLite s názvom construction.db (podobný súboru Excel s pracovnými listami), ktorý bude obsahovať tabuľky "Structural Impact" (vplyv na konštrukciu) a "Installation Schedule" (plán inštalácie).

Po vytvorení novej databázy **construction.db** v týchto tabuľkách je možné do týchto tabuľiek pridávať údaje prostredníctvom SQL \_dotazov alebo importu, čo vám neskôr umožní začať vytvárať automatické spracovanie údajov. Údaje možno do databázy SQLite importovať zo súborov CSV, z tabuľiek Excelu alebo exportovať z iných databáz a úložísk prostredníctvom rozhrania API.

Na vytvorenie udržateľného modelovania údajov a účinných procesov správy databáz potrebuje spoločnosť jasne definovanú strategiu, ako aj koordináciu medzi technickými a obchodnými tímmi. Pri rôznorodých projektoch a viacerých zdrojoch údajov je často ťažké zabezpečiť konzistentnosť, štandardizáciu a kontrolu kvality na všetkých úrovniach. Jedným z kľúčových riešení môže byť vytvorenie špecializovaného centra excellentnosti pre modelovanie údajov (CoE) v rámci spoločnosti.

## Centrum excellentnosti (CoE) pre modelovanie údajov

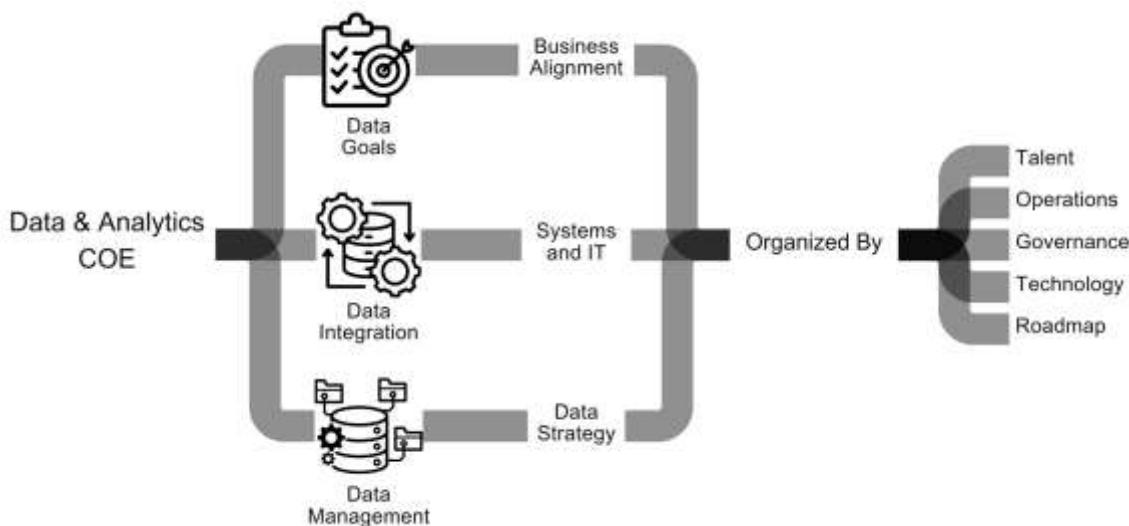
Kedže údaje sa stávajú jedným z kľúčových strategických aktív, spoločnosti musia robiť viac než len správne zhromažďovať a uchovávať informácie - je dôležité naučiť sa systematicky spravovať údaje. Centrum excellentnosti pre klasifikáciu a modelovanie údajov (CoE) je štrukturálna jednotka, ktorá zabezpečuje konzistentnosť, kvalitu a efektívnosť všetkého nakladania s údajmi v organizácii.

Centrum excellentnosti (CoE) je jadrom odbornej podpory a metodickým základom digitálnej transformácie v podniku. Buduje kultúru založenú na údajoch a umožňuje organizáciám budovať procesy, ktoré prijímajú rozhodnutia na základe štruktúrovaných, overených a reprezentatívnych údajov, a nie na základe intuúcie alebo lokálnych informácií.

Dátové centrum excellentnosti je zvyčajne vytvorené z multifunkčných tímov, ktoré pracujú na princípe "dvoch pizz". Tento princíp, ktorý navrhol Jeff Bezos, znamená, že veľkosť tímu by mala byť taká, aby sa dokázal nasýtiť dvoma pizzami, t. j. nie viac ako 6 až 10 ľudí. Tento prístup pomáha vyhnúť sa nadmernej byrokracie a zvyšuje flexibilitu práce. Tím CoE by mal zahŕňať zamestnancov s rôznymi technickými zručnosťami, od dátovej analýzy a strojového učenia až po odborné znalosti v konkrétnych obchodných oblastiach. Vďaka svojim hlbokým technickým znalostiam by mali dátoví inžinieri nielen optimalizovať procesy a modelovať údaje, ale aj podporovať kolegov tým, že skrátia čas na rutinné úlohy (obr. 4.3-9).

Tak ako v prírode odolnosť ekosystému zabezpečuje biodiverzita, v digitálnom svete sa flexibilita a prispôsobivosť dosahuje prostredníctvom rozmanitosti prístupov k spracovaniu údajov. Táto rozmanitosť však musí byť podporená spoločnými pravidlami a koncepciami.

Centrum excelentnosti (CoE) možno pirovnať ku "klimatickým podmienkam" lesného ekosystému, ktoré určujú, ktoré typy údajov budú prosperovať a ktoré budú automaticky vyradené. Vytvorením priaznivej "klímy" pre kvalitné údaje CoE uľahčuje prirodzený výber osvedčených postupov a metodík, ktoré sa potom stávajú štandardmi organizácie.



Obr. 4.3-9 Centrum excelentnosti (CoE) pre údaje a analýzu spája odborné znalosti o kľúčových aspektoch správy, integrácie a stratégie údajov.

V záujme urýchlenia integračných cyklov a dosiahnutia lepších výsledkov by mal CoE poskytnúť svojim členom dostatočnú mieru autonómie pri rozhodovaní. Je to dôležité najmä v dynamickom prostredí, kde pokusy a omyly, neustála spätná väzba a časté vydávanie verzií môžu priniesť významné výhody. Táto autonómia je však účinná len vtedy, ak existuje jasná komunikácia a podpora zo strany vyššieho manažmentu. Bez strategickej vízie a koordinácie na najvyššej úrovni môže aj ten najkompetentnejší tím čeliť prekážkam pri realizácii svojich iniciatív.

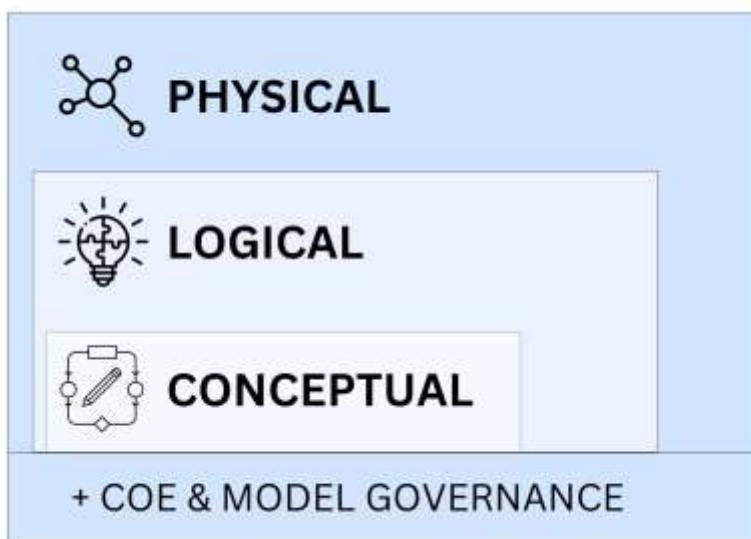
Práve COE alebo vrcholový manažment spoločnosti je zodpovedný za to, aby sa prístup k modelovaniu údajov neobmedzil na jeden alebo dva projekty, ale aby bol začlenený do celkového systému riadenia informácií a obchodných procesov.

Centrum expertízy (CoE) je okrem úloh súvisiacich s modelovaním údajov a správou údajov zodpovedné za vývoj spoločných noriem a prístupov k zavádzaniu a prevádzke dátovej infraštruktúry. Okrem toho podporuje kultúru neustáleho zlepšovania, optimalizáciu procesov a efektívne využívanie údajov v organizácii (obrázok 4.3-10).

Systematický prístup k správe údajov a modelov v rámci CoE možno zhruba rozdeliť do niekoľkých kľúčových blokov:

- **Štandardizácia procesov a riadenie životného cyklu modelov:** CoE vyvíja a zavádzza metodiky na zjednotenie tvorby a riadenia dátových modelov. To zahŕňa: vytvorenie štrukturálnych šablón, metód kontroly kvality a systémov kontroly verzií na zabezpečenie kontinuity údajov vo všetkých fázach práce.

- **Riadenie úloh a prideľovanie zodpovednosti:** CoE definuje kľúčové úlohy v procese modelovania údajov. Každému účastníkovi projektu sú pridelené jasne definované úlohy a oblasti zodpovednosti, čo uľahčuje tímovú prácu a znižuje riziko nekonzistentnosti údajov.
- **Kontrola kvality a audit:** efektívne riadenie údajov o stavbe si vyžaduje priebežné monitorovanie ich kvality. Zavádzajú sa automatizované mechanizmy na kontrolu údajov, identifikáciu chýb, chýbajúcich atribútov.
- **Metadáta a riadenie informačnej architektúry:** CoE je zodpovedný za vytvorenie jednotného systému klasifikácie a identifikátorov, štandardov pre pomenovanie a popis entít, ktorý je rozhodujúci pre integráciu medzi systémami.



Obr. 4.3-10 Modelovanie údajov a riadenie kvality údajov je jednou z hlavných výziev CoE

Centrum excelentnosti (CoE) pre údaje nie je len skupina odborníkov, ale systémový mechanizmus, ktorý vytvára novú kultúru založenú na údajoch a zabezpečuje jednotný prístup k práci s údajmi v celej spoločnosti. Prostredníctvom kompetentnej integrácie procesov modelovania do celkového systému riadenia informácií, štandardizácie, klasifikácie a kontroly kvality údajov, CoE pomáha podnikom neustále zlepšovať svoje produkty a obchodné procesy, rýchlejšie reagovať na zmeny na trhu a prijímať informované rozhodnutia na základe spoločenských analýz.

Takéto centrá sú obzvlášť efektívne v kombinácii s modernými princípmi DataOps - pod krok, ktorý zabezpečuje kontinuálne poskytovanie, automatizáciu a kontrolu kvality údajov. O DataOps budeme viac hovoriť v 8. časti, v kapitole "Moderné dátové technológie v stavebníctve".

V nasledujúcich kapitolách prejdeme od stratégie k praxi - podmienečne sa "transformujme" do dátového centra: pozrieme sa na niekoľko príkladov, ako prebieha parametrizácia úloh, zber požiadaviek a proces automatickej validácie.



## KAPITOLA 4.4.

### SYSTEMATIZÁCIA POŽIADAVIEK A OVEROVANIE INFORMÁCIÍ

#### Zhromažďovanie a analýza požiadaviek: transformácia komunikácie na štruktúrované údaje

Zhromažďovanie a správa požiadaviek je prvým krokom na zabezpečenie kvality údajov. Napriek rozvoju digitálnych nástrojov sa väčšina požiadaviek stále formuluje neštruktúrovaným spôsobom: prostredníctvom listov, zápisníc zo stretnutí, telefonátov a ústnych diskusí. Táto forma komunikácie stáže automatizáciu, validáciu a opakované použitie informácií. V tejto kapitole sa zaoberáme tým, ako previesť textové požiadavky do formálnych štruktúr, čím sa zabezpečí, že obchodné požiadavky budú transparentné a systematické.

Výskum spoločnosti Gartner "Kvalita údajov: osvedčené postupy pre presné poznatky" zdôrazňuje kľúčový význam kvality údajov pre úspešné dátové a analytické iniciatívy [100]. Konštatujú, že zlá kvalita údajov stojí organizácie v priemere najmenej 12,9 milióna dolárov ročne a že spoľahlivé a kvalitné údaje sú nevyhnutné na vytvorenie spoločnosti založenej na údajoch.

Absencia štruktúrovaných požiadaviek vedie k tomu, že ten istý prvak (entita) a jeho parametre môžu byť uložené v rôznych systémoch v rôznych variantoch. To nielenže znížuje efektívnosť procesov, ale vedie aj k plynaniu časom, duplicitne informácií a potrebe opäťovného overovania údajov pred ich použitím. V dôsledku toho môže aj jediné opomenutie - stratený parameter alebo jedený nesprávne opísaný prvak - spomalit' rozhodovanie a spôsobiť neefektívne využívanie zdrojov.

*Pre nedostatok klinca sa podkova stratila.  
Pre nedostatok podkovy sa kôň stratil.  
Pre nedostatok koňa sa stratil jazdec.  
Pre nedostatok jazdca sa stratila správa.  
Pre nedostatok posolstva bola bitka prehratá.  
Pre nedostatok bitky bolo stratené kráľovstvo.  
Všetko pre nedostatok klinca v podkove.*

- Príslovie [101]

Analýza a zhromažďovanie požiadaviek na proces vypĺňania a ukladania údajov sa začína identifikáciou všetkých zainteresovaných strán. Tak ako príslovečná strata jediného klinca vedie k reťazi kritických dôsledkov, v podnikaní môže strata jedinej zainteresovanej strany, prehliadnutá požiadavka alebo strata čo i len jediného parametra výrazne ovplyvníť nielen jednotlivý podnikový proces, ale aj celý ekosystém projektu a organizácie ako celku. Preto je veľmi dôležité identifikovať aj tie prvky, parametre a úlohy, ktoré sa na prvý pohľad zdajú bezvýznamné, ale neskôr sa môžu ukázať ako kritické pre udržateľnosť podniku.

Predstavme si, že spoločnosť má projekt, v ktorom klient predloží novú požiadavku - "pridať ďalšie okno na severnej strane budovy". Na malom procese "požiadavka klienta na pridanie nového okna do súčasného projektu" sa podieľajú architekt, klient, špecialista CAD (BIM), stavebný manažér, manažér logistiky, ERP -analytik, inžinier kontroly kvality, bezpečnostný inžinier, manažér kontroly a manažér majetku.

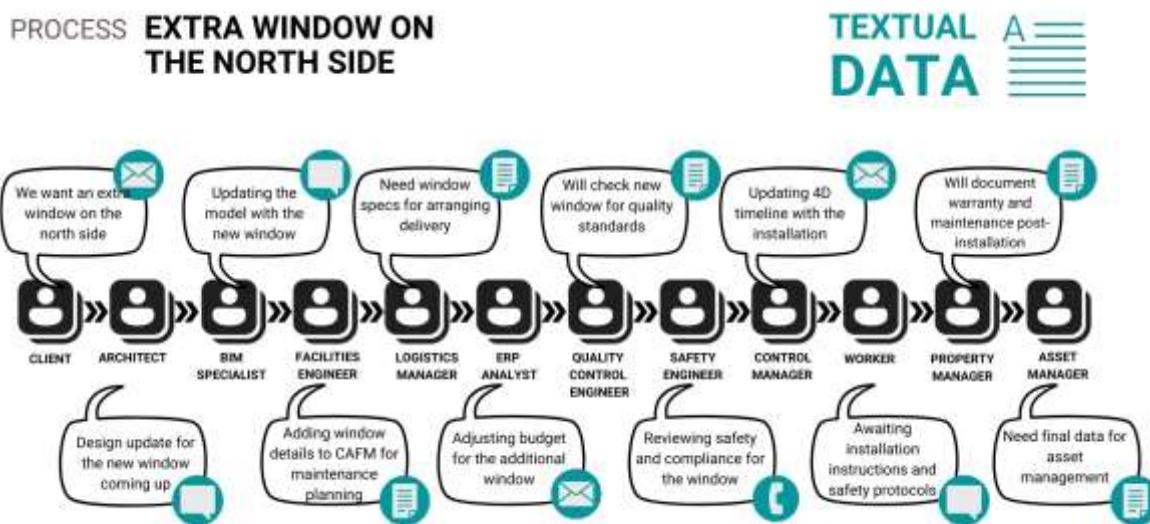
Aj na malom procese sa môžu podieľať desiatky rôznych odborníkov. Každý účastník procesu musí rozumieť požiadavkám špecialistov, s ktorými je prepojený na úrovni údajov.

Na úrovni textu (obr. 4.4-1) prebieha komunikácia medzi klientom a špecialistami v procesnom reťazci nasledovne:

- ⦿ **Zákazník:** "Rozhodli sme sa pridať ďalšie okno na severnej strane pre lepšie osvetlenie. Dá sa to zrealizovať?"
- ⦿ **Architekt:** "Iste, prepracujem projekt tak, aby obsahoval nové okno, a pošlem aktualizované plány CAD (BIM)".
- ⦿ **Špecialista CAD (BIM):** "Priyatý nový projekt. Aktualizujem model CAD (BIM) o ďalšie okno a po koordinácii s inžinierom FEM poskytnem presné umiestnenie a rozmery nového okna".
- ⦿ **Vedúci výstavby:** "Bol prijatý nový projekt. Upravujeme termíny inštalácie 4D a informujeme všetkých príslušných subdodávateľov."
- ⦿ **Inžinier zariadení (CAFM):** "Údaje 6D na novom okne vložím do systému CAFM pre budúce plánovanie riadenia a údržby zariadení."
- ⦿ **Manažér logistiky:** "Potrebujem rozmery a hmotnosť nového okna, aby som mohol zorganizovať jeho dodávku na stavbu."
- ⦿ **ERP -analytik:** "Potrebujem tabuľky rozsahu a presný typ okna pre aktualizáciu rozpočtu 5D v našom systéme ERP, aby sa náklady na nové okno premietli do celkového odhadu projektu."
- ⦿ **Inžinier kontroly kvality:** "Keď budú špecifikácie okien hotové, zabezpečím, aby splňali naše normy kvality a materiálu."
- ⦿ **Bezpečnostný inžinier:** "Budem posudzovať bezpečnostné aspekty nového okna, pričom sa zameřiam najmä na dodržiavanie predpisov a evakuáciu podľa schémy 8D".
- ⦿ **Manažér kontroly:** "Na základe presného rozsahu prác z ERP aktualizujeme našu časovú os 4D tak, aby odrážala inštaláciu nového okna, a uložíme nové údaje do systému správy obsahu projektu."
- ⦿ **Pracovník (inštalatér):** "Potrebujete pokyny na inštaláciu, montáž a načasovanie práce. Okrem toho, sú zavedené nejaké špeciálne bezpečnostné pravidlá, ktoré musím dodržiavať?"
- ⦿ **Správca nehnuteľnosti:** "Po inštalácii zdokumentujem informácie o záruke a údržbe pre dlhodobú správu".
- ⦿ **Správca majetku:** "Inžinier zariadení, pošlite, prosím, konečné údaje pre sledovanie a riadenie životného cyklu majetku."
- ⦿ **Klient:** "Počkajte, možno sa ponáhľam a okno nebude potrebné. Možno by som mal urobiť balkón."

V takýchto scenároch, ktoré sa vyskytujú často, spôsobí aj malá zmena reťazovú reakciu medzi viacerými systémami a rolami. V tomto prípade je takmer všetka komunikácia v počiatočnej fáze v textovej forme: e-maily, chaty, zápisu zo stretnutí (obr. 4.4-1).

V takomto systéme textovej komunikácie pre stavebný projekt je veľmi dôležitý systém právneho potvrdenia a zaznamenávania všetkých operácií výmeny údajov a všetkých prijatých rozhodnutí. Tým sa zabezpečí, aby každé rozhodnutie, pokyn alebo vykonaná zmena boli právne platné a vysledovateľné, čím sa zníži riziko budúcich "nedorozumení".

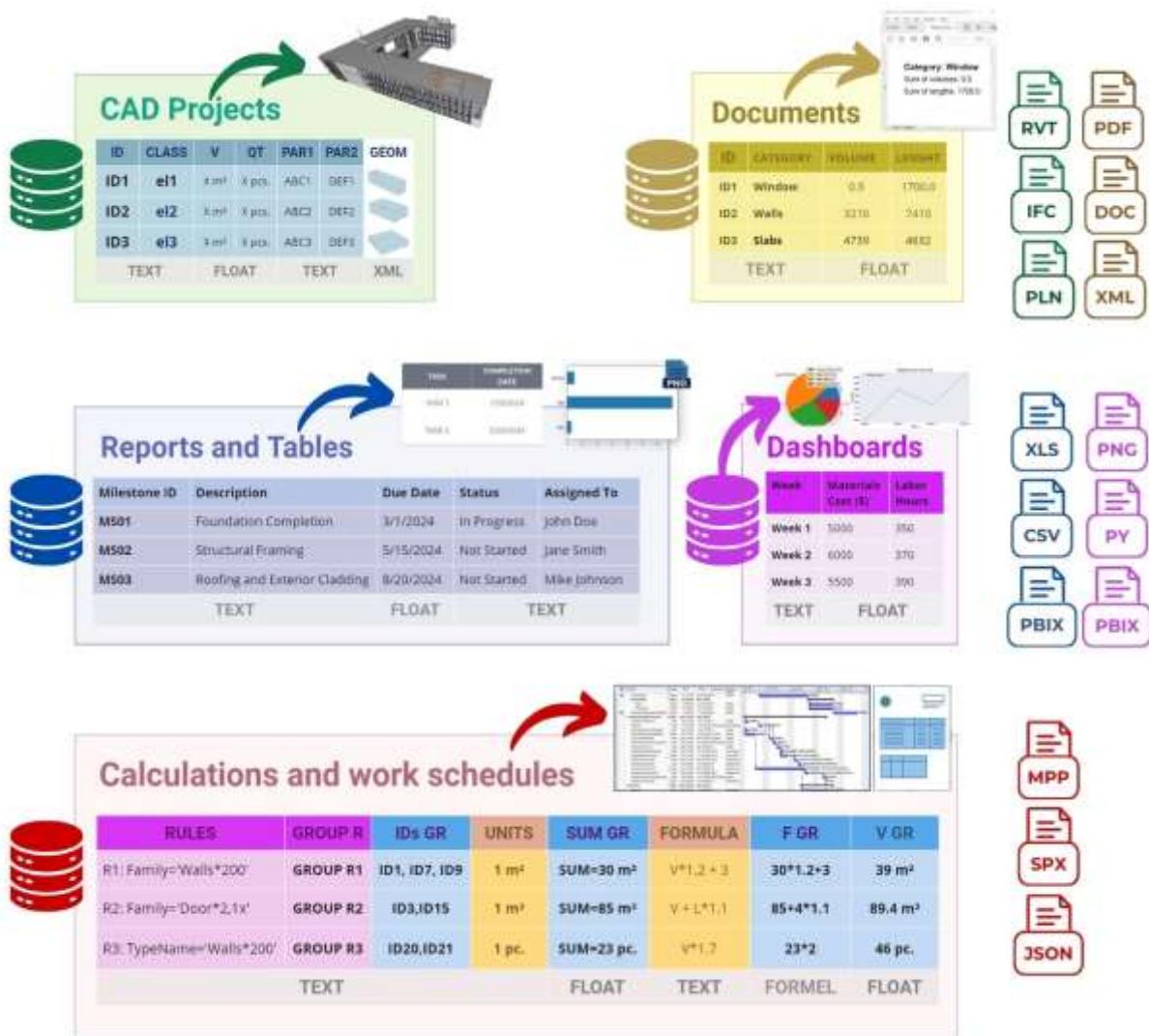


Obr. 4.4-1 Komunikácia medzi klientom a dodávateľom v počiatocných fázach projektu často obsahuje textové údaje vo viacerých formátoch.

Nedostatočná právna kontrola a validácia rozhodnutí v príslušných systémoch stavebného projektu môže viesť k vážnym problémom pre všetkých zúčastnených. Každé rozhodnutie, príkaz alebo zmena vykonaná bez riadnej dokumentácie a validácie môže viesť k sporom (a súdnym sporom).

Právnu konsolidáciu všetkých rozhodnutí v textovej komunikácii možno zabezpečiť len veľkým počtom podpísaných dokumentov, ktoré budú ležať na pleciach manažmentu, ktorý je povinný zaznamenávať všetky transakcie. V dôsledku toho, ak každý účastník musí podpísovať dokumenty pre každý úkon, systém stráca flexibilitu a stáva sa byrokratickým labyrintom. Nedostatok potvrdení o transakciách nielenže oneskorí realizáciu projektu, ale môže viesť aj k finančným stratám a zhoršeniu vzťahov medzi účastníkmi až po právne problémy.

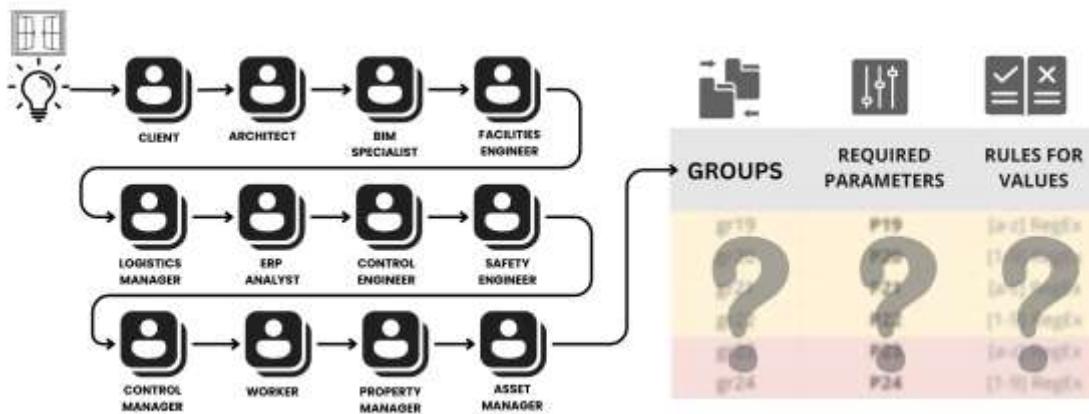
Takýto proces schvaľovania transakcií, ktorý sa zvyčajne začína textovou diskusiou, sa v nasledujúcich fázach postupne mení na výmenu dokumentov vo viacerých formátoch (obr. 4.4-2), čím sa výrazne komplikuje komunikácia, ktorá predtým prebiehala len prostredníctvom textu. Bez jasne definovaných požiadaviek sa automatizácia takýchto procesov, ktoré sú plné multiformátových údajov a veľkého množstva textových požiadaviek, stáva takmer nemožnou.



Obr. 4.4-2 Každý systém v prostredí stavebnej firmy slúži ako zdroj právne relevantných dokumentov v rôznych formátoch.

Textová komunikácia si vyžaduje, aby sa každý odborník buď oboznámil s celou korešpondenciou, alebo sa pravidelne zúčastňoval všetkých stretnutí, aby pochopil aktuálny stav projektu.

Na prekonanie tohto obmedzenia je potrebný prechod od textovej komunikácie k štruktúrovanému modelu požiadaviek. To je možné len prostredníctvom systematickej analýzy, vizualizácie procesov a opisu interakcií vo forme vývojových diagramov a dátových modelov (obr. 4.4-3). Rovnako ako pri dátovom modelovaní (obr. 4.3-7) sme sa z kontextovo-ideovej úrovne presunuli na konceptuálnu úroveň pridaním systémov a nástrojov, ktoré účastníci používajú, a väzieb medzi nimi.



Obr. 4.4-3 Aby sme sa naučili riadiť a automatizovať proces validácie, je potrebné vizualizovať procesy a štruktúrovať požiadavky.

Prvým krokom pri systematizácii požiadaviek a vzťahov je vizualizácia všetkých väzieb a vzťahov pomocou konceptuálnych vývojových diagramov. Konceptuálna úroveň nielen uľahčí všetkým účastníkom procesu pochopiť celý reťazec procesov, ale tiež jasne ukáže, prečo a pre koho sú údaje (a požiadavky) potrebné v každom kroku procesu.

### Procesné diagramy a účinnosť koncepčných rámcov

Na preklenutie prienosti medzi tradičným a moderným prístupom k správe údajov musia spoločnosti vedome prejsť od fragmentovaných textových opisov k štruktúrovaným reprezentáciám procesov. Vývoj údajov - od hlinených tabuľiek k digitálnym ekosystémom - si vyžaduje nové nástroje myslenia. A jedným z takýchto nástrojov je konceptuálne modelovanie pomocou vývojových diagramov. Vytváranie vizuálnych diagramov - vývojových diagramov, procesných diagramov, interakčných diagramov - umožňuje účastníkom projektu uvedomiť si, ako ich činnosti a rozhodnutia ovplyvňujú celý systém rozhodovania.

Ak si procesy vyžadujú nielen ukladanie údajov, ale aj ich analýzu alebo automatizáciu, musíte sa začať zaoberať tému vytvorenia koncepčnej a vizuálnej vrstvy požiadaviek.

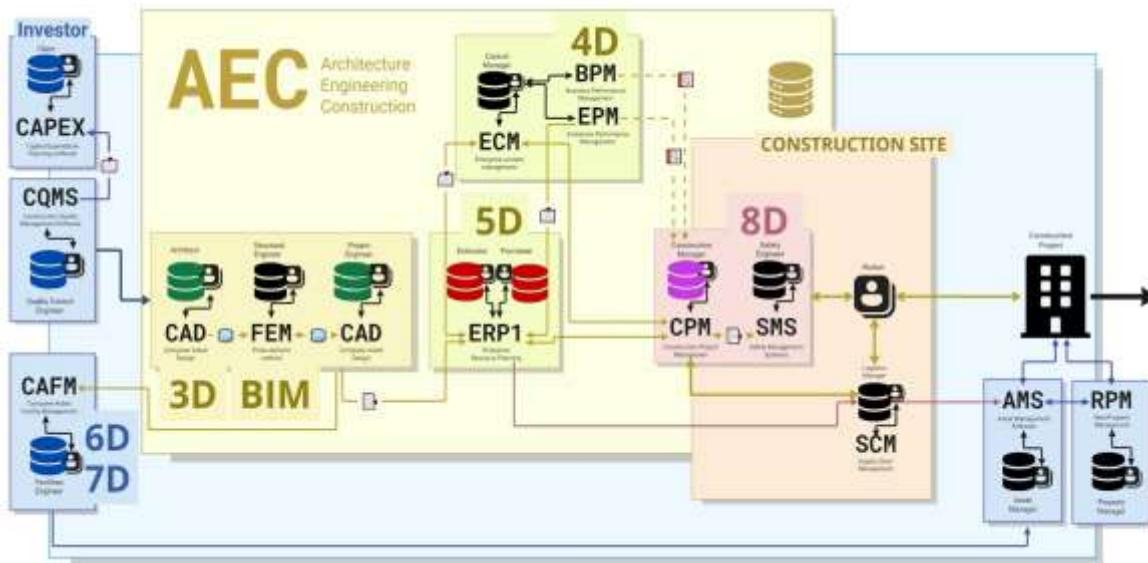
V našom príklade (obr. 4.4-1) môže byť každý špecialista súčasťou malého tímu, ale aj väčšieho oddelenia s až desiatimi odborníkmi pod kontrolou generálneho manažéra. Každé oddelenie používa špecializovanú aplikačnú databázu (obr. 1.2-4, napr. ERP, CAD, MEP, CDE, ECM, CPM atď.), ktorá sa pravidelne aktualizuje o prichádzajúce informácie potrebné na tvorbu dokumentov, zaznamenávanie právneho stavu rozhodnutí a riadenie procesov.

Proces transakcie je podobný práci starovekých manažérov spred 4 000 rokov, keď sa na právne potvrdenie rozhodnutí používali hlinené tabuľky a papyrus. Rozdiel medzi modernými systémami a ich hlinenými a papierovými predchodcami je v tom, že moderne metódy navyše zahŕňajú proces prevodu textových informácií do digitálnej podoby na ďalšie automatické spracovanie v iných systémoch a nástrojoch.

Vytvorenie vizualizácie procesu vo forme koncepcných vývojových diagramov pomôže opísť každý krok a interakcie medzi jednotlivými úlohami, čím sa zložitý pracovný postup stane jasným a jednoduchým.

Vizualizácia procesov zabezpečuje, že logika procesov je transparentná a prístupná všetkým členom tímu.

Rovnaký komunikačný proces pridávania okna do projektu, ktorý bol opísaný vo forme textu, správ (obrázok 4.4-1) a blokového diagramu, je podobný koncepcnému modelu, o ktorom sme hovorili v kapitole o modelovaní údajov (obrázok 4.4-4).



Obr. 4.4-4 Koncepcný diagram zobrazuje účastníkov projektu ako používateľov databáz, kde ich dotazy spájajú rôzne systémy.

Hoci sú koncepcné diagrame dôležitým krokom, mnohé spoločnosti sa obmedzujú na túto úroveň a veria, že vizuálny diagram je dostatočný na pochopenie procesov. Vytvára sa tak ilúzia zvládnuteľnosti: manažéri môžu na takomto vývojovom diagrame ľahšie vnímať celkový obraz a vidieť súvislosti medzi účastníkmi a etapami. Takéto schémy však neposkytujú jasné predstavu o tom, aké údaje sú potrebné pre každého účastníka, v akom formáte sa majú prenášať a aké parametre a atribúty sú povinné na realizáciu automatizácie. Konceptuálna vývojová schéma je skôr ako mapa trasy: naznačuje, kto s kým komunikuje, ale neodhaľuje, čo sa v týchto interakciách prenáša.

Aj keď je proces podrobne opísaný na koncepcnej úrovni pomocou vývojových diagramov, nezaručuje to jeho účinnosť. Vizualizácia často zjednodušuje prácu manažérov a umožňuje im ľahšie sledovať proces pomocou krokových správ. Pre databázových inžinierov však nemusí byť konceptuálne znázornenie dostatočne jasné a nemusí poskytovať jasné pochopenie toho, ako proces implementovať na úrovni parametrov a požiadaviek.

S prechodom na komplexnejšie dátové ekosystémy sa počiatočná implementácia koncepcných a vizuálnych nástrojov stáva rozhodujúcou, aby sa zabezpečilo, že dátové procesy budú nielen efektívne, ale aj v súlade so strategickými cieľmi organizácie. Aby sme tento proces pridávania okien (obrázok 4.4-1) plne preniesli na úroveň požiadaviek na údaje, musíme ísť o úroveň hlbšie a previesť koncepcnú vizualizáciu procesu na logickú a fyzickú úroveň údajov, požadovaných atribútov a ich hraničných hodnôt.

## Štruktúrované požiadavky a regulárne výrazy RegEx

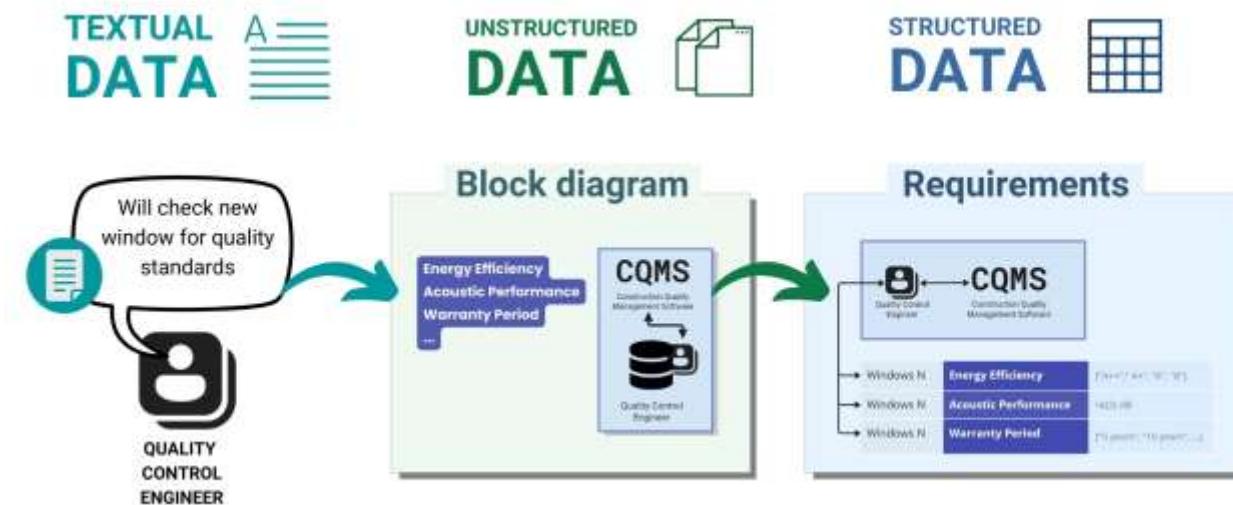
Až 80% údajov vytvorených vo firmách je v neštruktúrovaných alebo pološtruktúrovaných formátoch [52] - text, dokumenty, listy, PDF - súbory, rozhovory. Takéto údaje (obr. 4.4-1) sa ľahko analyzujú, overujú, prenášajú medzi systémami a využívajú v automatizácii.

Na zabezpečenie spravovateľnosti, transparentnosti a automatickej validácie je potrebné previesť textové a pološtruktúrované požiadavky do presne definovaných štruktúrovaných formátov. Proces štruktúrovania sa týka nielen údajov (ktorým sme sa podrobne venovali v prvých kapitolách tejto časti knihy), ale aj samotných požiadaviek, ktoré účastníci projektu zvyčajne formulujú vo voľnej textovej podobe počas celého životného cyklu projektu, často bez toho, aby mysleli na to, že tieto procesy možno automatizovať.

Tak ako sme už previedli údaje z neštruktúrovanej textovej formy do štruktúrovanej formy, v pracovnom postupe požiadaviek prevedieme textové požiadavky do štruktúrovaného formátu "logickej a fyzickej vrstvy".

V rámci príkladu pridania okna (obrázok 4.4-1) je ďalším krokom opis požiadaviek na údaje v tabuľkovej forme. Informácie pre každý systém používaný účastníkmi projektu budeme štruktúrovať uvedením kľúčových atribútov a ich hraničných hodnôt

Zoberme si napríklad jeden takýto systém (obr. 4.4-5) - systém riadenia kvality stavby (CQMS), ktorý používa inžinier kontroly kvality na strane zákazníka. S jeho pomocou kontroluje, či nový prvk projektu - v tomto prípade "nové okno" - zodpovedá stanoveným normám a požiadavkám.



Obr. 4.4-5 Konverzia textových požiadaviek do tabuľkového formátu s popisom atribútov entity zjednodušuje pochopenie pre ostatných špecialistov.

Ako príklad uveďme niektoré dôležité požiadavky na atribúty entít typu "okenné systémy" v systéme CQMS (obr. 4.4-6): energetická účinnosť, akustické vlastnosti a záručná doba. Každá kategória zahŕňa určité normy a špecifikácie, ktoré je potrebné zohľadniť pri navrhovaní a montáži okenných systémov.



Obr. 4.4-6 Inžinier kontroly kvality by mal skontrolovať nové prvky typu Window z hľadiska energetickej účinnosti, zvukovej izolácie a záručných noriem.

Požiadavky na údaje, ktoré technik kontroly kvality špecifikuje vo forme tabuľky, majú napríklad tieto hraničné hodnoty:

- **Trieda energetickej účinnosti okien** sa pohybuje od "A++", ktorá označuje najvyššiu účinnosť, po "B", ktorá sa považuje za minimálnu prijateľnú úroveň, a tieto triedy sú reprezentované zoznamom prijateľných hodnôt **["A++", "A+", "A", "A", "A", "B"]**.
- **Akustická izolácia okien**, meraná v decibeloch a vyjadrujúca ich schopnosť znižovať hluk z

ulice, je definovaná pravidelným výrazom \d{2}dB.

- **Atribút "Záručná doba"** pre entitu "Typ okna" začína na piatich rokoch, čím sa táto doba nastavuje ako minimálna povolená pri výbere výrobku; zadávajú sa aj hodnoty záručnej doby, ako napríklad **"5 rokov", "10 rokov" atď.** alebo logická podmienka **">5 (rokov)"**.

Podľa zozbieraných požiadaviek v rámci stanovených atribútov nové prvky kategórie alebo triedy okien so známkou nižšou ako "B", napríklad "C" alebo "D", neprejdú testom energetickej účinnosti. Akustická izolácia okien v údajoch alebo dokumentoch pre inžiniera zabezpečenia kvality sa označuje dvojmiestnym číslom, za ktorým nasleduje postfix "dB", napríklad "35 dB" alebo "40 dB", a hodnoty mimo tohto formátu, ako napríklad "9 D B" alebo "100 decibelov", nebudú akceptované (pretože neprejdú vzorom pre reťazce RegEx). Záručná doba musí začínať minimálne "5 rokov" a okná s kratšími záručnými dobami, ako napríklad "3 roky" alebo "4 roky", nebudú spĺňať požiadavky, ktoré inžinier kvality opísal vo formáte tabuľky.

Na kontrolu takýchto hodnôt atribútov-parametrov voči hraničným hodnotám z požiadaviek v procese validácie používame buď zoznam povolených hodnôt ([A", "B", "C"]), slovníky ([A": "H1", "H2"; "B": W1", "W2"]), logické operácie (napr. ">", "<", "<=", ">" "=") pre číselné hodnoty) a regulárne výrazy (pre reťazcové a textové hodnoty, ako napríklad v atribúte "Acoustic Performance"). Regulárne výrazy sú mimoriadne dôležitým nástrojom pri práci s reťazcovými hodnotami.

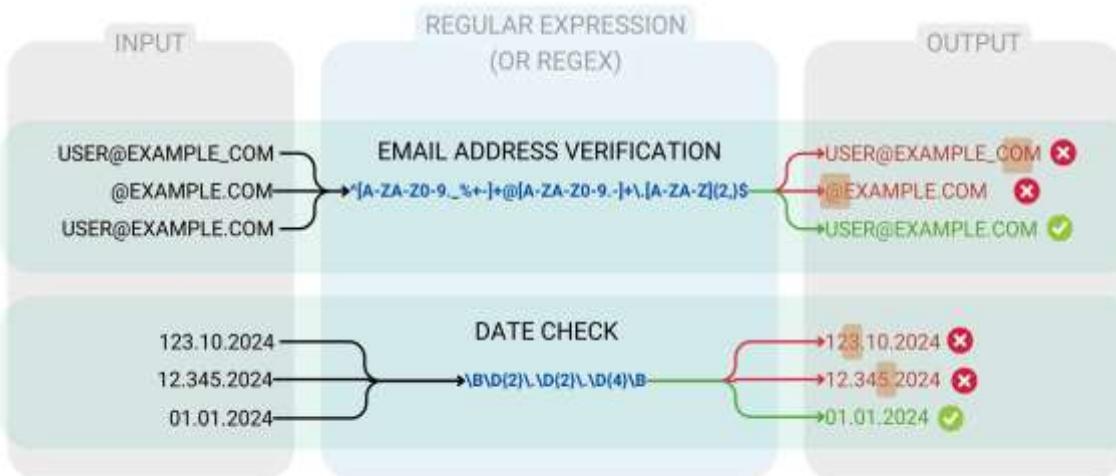
**Regulárne výrazy** (RegEx) sa používajú v programovacích jazykoch vrátane jazyka Python (knižnica Re) na vyhľadávanie a úpravu reťazcov. Regex je ako detektív vo svete reťazcov, ktorý dokáže presne identifikovať textové vzory v teste.

V regulárnych výrazoch sa písmená opisujú priamo pomocou príslušných znakov abecedy, zatiaľ čo čísla možno reprezentovať pomocou špeciálneho znaku \d, ktorý zodpovedá akékoľvek číslici od 0 do 9. Hranaté zátvorky sa používajú na označenie rozsahu písmen alebo číslíc, napr. [a-z] pre akékoľvek malé písmeno latinskej abecedy alebo [0-9], ktoré zodpovedá \d. Pre nečíselné a nepísmenové znaky sa používa \D a \W.

Oblúbené prípady použitia RegEx (obrázok 4.4-7):

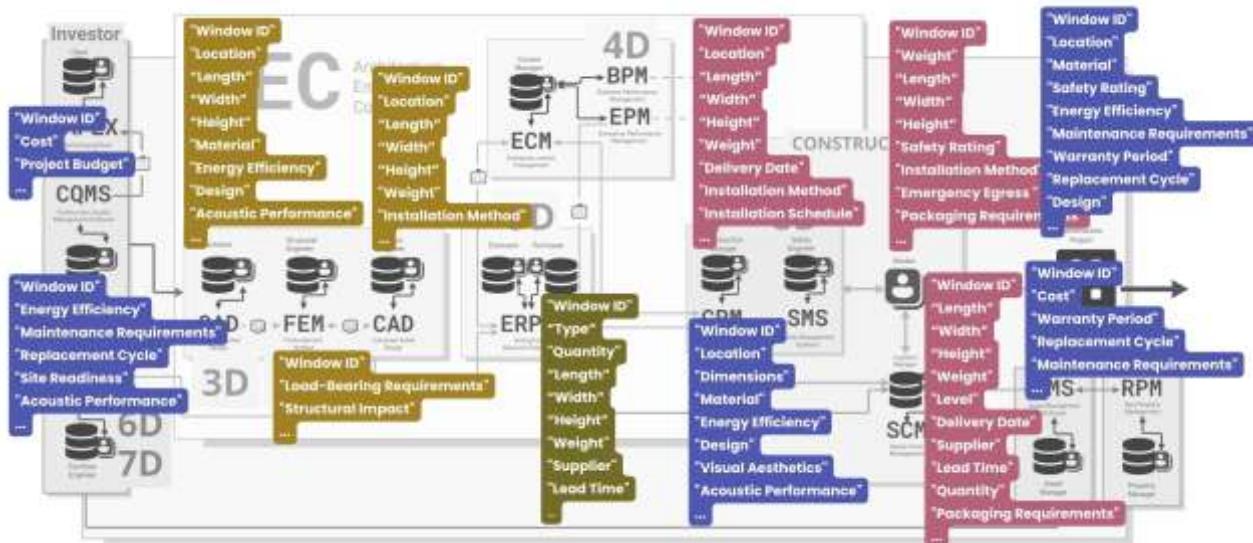
- **Overenie e-mailovej adresy** : Ak chcete overiť, či je reťazec platnou e-mailovou adresou, môžete použiť šablónu "**^ [a-zA-Z0-9.\_%+-]+@[a-zA-Z0-9.-]+\.[a-zA-Z]{2,}\$**".
- **Extrakcia dátumu**: Šablónu "**\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b**" možno použiť na extrakciu dátumu z textu vo formáte DD.MM.RRRR.
- **Overenie telefónnych čísel** : Ak chcete overiť telefónne čísla vo formáte +49(000)000-0000, vzor bude vyzeráť takto: "**\+\d{2}\(\d{3}\)\d{3}-\d{4}**".

Prevodom požiadaviek inžiniera kvality do formátu atribútov a ich hraničných hodnôt (obr. 4.4-6) sme ich z pôvodného textového formátu (rozhovory, listy a regulačné dokumenty) transformovali do usporiadanej a štruktúrovanej tabuľky, čím sme umožnili automatickú kontrolu a analýzu všetkých prichádzajúcich údajov (napr. nových prvkov kategórie Okno). Prítomnosť požiadaviek umožňuje automaticky vyradiť údaje, ktoré neboli skontrolované, a skontrolované údaje automaticky preniesť do systémov na ďalšie spracovanie.



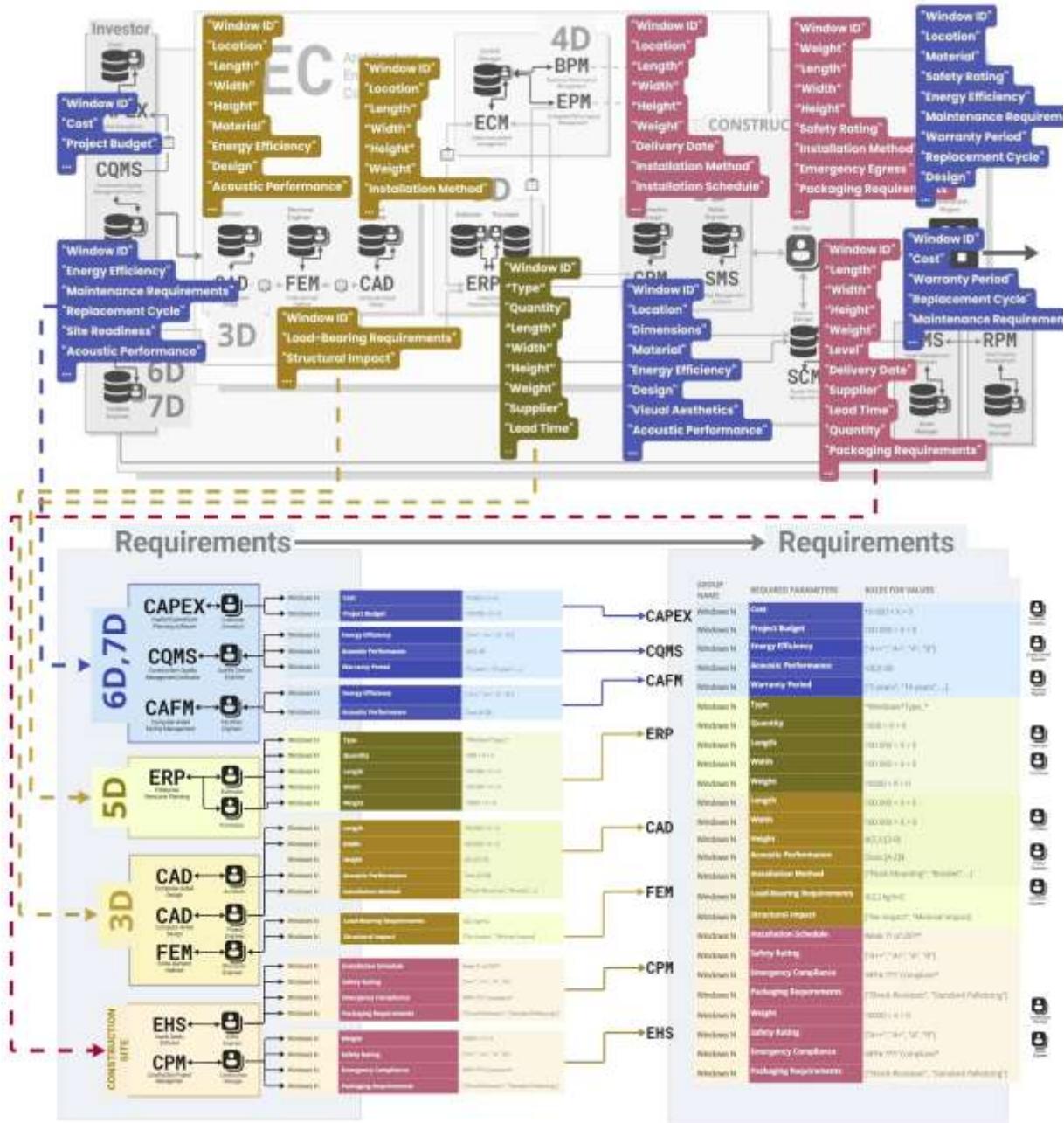
Obrázok 4.4-7 Používanie regulárnych výrazov je mimoriadne dôležitým nástrojom v procese overovania textových údajov.

Teraz, keď prejdeme z konceptuálnej na logickú úroveň práce s požiadavkami, prevedieme všetky požiadavky všetkých špecialistov v našom procese inštalácie nového okna (obr. 4.4-4) na usporiadany zoznam vo formáte atribútov a tieto zoznamy s potrebnými atribútmi pridáme do nášho vývojového diagramu pre každého špecialistu (obr. 4.4-8).



Obrázok 4.4-8 Na úrovni logických procesov sa atribúty, ktoré spracúvajú jednotliví špecialisti, pridávajú do ich príslušných systémov.

Pridaním všetkých atribútov do jednej spoločnej procesnej tabuľky transformujeme informácie, ktoré boli predtým prezentované ako text a dialóg na konceptuálnej úrovni (obr. 4.4-1), do štruktúrovanej a systematickej podoby tabuľiek fyzickej úrovne (obr. 4.4-9).



Obrázok 4.4-9 Konverzia neštruktúrovaných odborných dialógov do štruktúrovaných tabuľiek pomáha pochopíť požiadavky na fyzickej úrovni.

Teraz je potrebné označiť požiadavky na údaje odborníkom, ktorí vytvárajú informácie pre konkrétné systémy. Ak napríklad pracujete v databáze CAD, pred začatím modelovania prvkov by ste mali zhromaždiť všetky potrebné parametre na základe scenárov konečného použitia údajov. Zvyčajne sa začína prevádzkovou fázou, po ktorej nasleduje stavenisko, oddelenie logistiky, oddelenie odhadov, oddelenie statických výpočtov atď. Až po zohľadnení požiadaviek všetkých týchto oblastí môžete začať vytvárať údaje - na základe zozbieraných parametrov. To vám v budúcnosti umožní automatizovať overovanie a prenos údajov v rámci reťazca.

Ked' nové údaje spĺňajú požiadavky, automaticky sa integrujú do dátového ekosystému spoločnosti a dostanú sa priamo k používateľom a systémom, pre ktoré boli určené. Overovanie údajov na základe atribútov a ich hodnôt zabezpečuje, že informácie spĺňajú požadované normy kvality a sú pripravené na použitie v podnikových scenároch.

Požiadavky na údaje boli definované a teraz, pred začatím overovania, je potrebné vytvoriť, získať alebo zhromaždiť údaje, ktoré sa majú overiť, alebo zaznamenať aktuálny stav informácií v databázach, aby sa mohli použiť v procese overovania.

## Zber údajov pre proces overovania

Pred začatím validácie je dôležité zabezpečiť, aby boli údaje k dispozícii vo forme vhodnej pre proces validácie. To neznamená len mať informácie k dispozícii, ale pripraviť ich: údaje sa musia zhromaždiť a transformovať z neštruktúrovaných, voľne štruktúrovaných, textových a geometrických formátov do štruktúrovanej podoby. Tento proces je podrobne opísaný v predchádzajúcich kapitolách, v ktorých sa rozoberali metódy transformácie rôznych typov údajov. Výsledkom všetkých transformácií je, že prichádzajúce údaje majú podobu otvorených štruktúrovaných tabuľiek (obr. 4.1-2, obr. 4.1-9, obr. 4.1-13).

S požiadavkami a štruktúrovanými tabuľkami s potrebnými parametrami a hraničnými hodnotami (obr. 4.4-9) môžeme začať validovať údaje - bud' ako jeden automatizovaný proces (Pipeline), alebo ako postupnú validáciu každého prijatého dokumentu.

Na spustenie kontroly je potrebné buď priať nový súbor ako vstup, alebo opraviť aktuálny stav údajov - vytvoriť snímku alebo exportovať aktuálne a prichádzajúce údaje, alebo nastaviť pripojenie k externej alebo internej databáze. V posudzovanom príklade sa takýto snímok vytvorí automatickou konverziou údajov CAD zo do štruktúrovaného formátu zaznamenaného napríklad v piatok 29. marca 2024 o 23:00:00 h po tom, ako všetci projektanti odišli domov.



Obrázok 4.4-10 Snímka databázy CAD (BIM) zobrazujúca aktuálne informácie o atribútoch novej entity triedy "Okno" v aktuálnej verzii modelu projektu.

Vďaka nástrojom reverzného inžinierstva, o ktorých sa hovorí v kapitole "Prevod údajov CAD (BIM) do štruktúrovanej podoby", možno tieto informácie z rôznych nástrojov CAD (BIM) a editorov usporiadať

do samostatných tabuľiek (obr. 4.4-11) alebo ich spojiť do jednej spoločnej tabuľky, ktorá spája rôzne časti projektu (obr. 9.1-10).

Takáto tabuľka - databáza zobrazuje jedinečné identifikátory okien a dverí (atribút ID), názvy typov (TypeName), rozmery (Width, Length), materiály (Material), ako aj ukazovatele energetických a akustických vlastností a ďalšie charakteristiky. Takúto tabuľku vyplnenú v programe CAD (BIM) zhromažďuje projektant z rôznych oddelení a dokumentov a vytvára informačný model projektu.



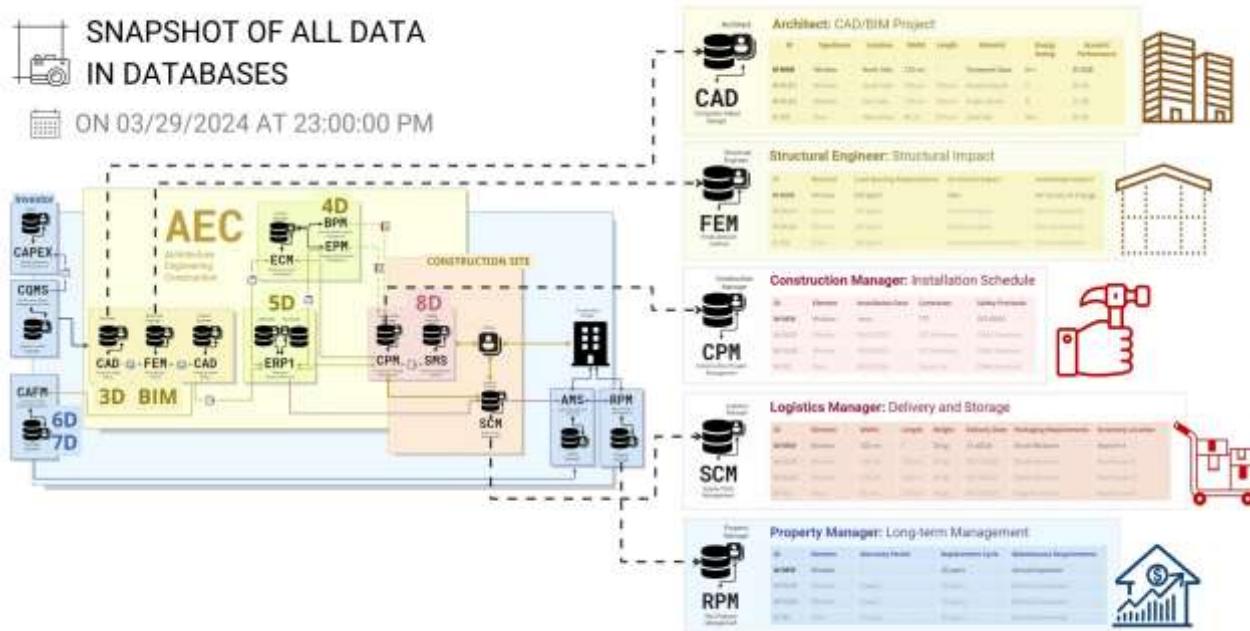
Obrázok 4.4-11 Štruktúrované údaje zo systémov CAD môžu mať podobu dvojrozmernej tabuľky so stĺpcami označujúcimi atribúty prvkov.

Reálne projekty CAD (BIM) obsahujú desiatky alebo stovky tisíc prvkov (obr. 9.1-10). Prvky v rámci formátov CAD sú automaticky kategorizované podľa typu a kategórie - od okien a dverí až po dosky, podlahy a steny. Jedinečné identifikátory (napr. natívne ID, ktoré automaticky nastavuje riešenie CAD) alebo atribúty typu (názov typu, typ, rodina) umožňujú sledovať ten istý objekt v rôznych systémoch. Napríklad nové okno na severnej stene budovy možno jednoznačne identifikovať pomocou jediného identifikátora "W-NEW" vo všetkých príslušných systémoch organizácie.

Zatial' čo názvy a identifikátory entít by mali byť konzistentné vo všetkých systémoch, súbor atribútov a hodnôt spojených s týmito entitami sa môže výrazne lísiť v závislosti od kontextu použitia. Architekti, stavební inžinieri, odborníci na výstavbu, logistiku a údržbu nehnuteľností vnímajú tie isté prvky rôznymi spôsobmi. Každý z nich sa spolieha na vlastné klasifikátory, normy a ciele: niektorí sa na okno pozerajú čisto z estetického hľadiska, posudzujú jeho tvar a proporcie, iní sa naň pozerajú z technického alebo prevádzkového hľadiska, analyzujú tepelnú vodivosť, spôsob inštalácie, hmotnosť alebo požiadavky na údržbu. Pri modelovaní údajov a popise prvkov je preto dôležité zohľadniť všeestrannosť ich použitia a zabezpečiť konzistentnosť údajov pri zohľadnení špecifík odvetvia.

Pre každú úlohu v procesoch spoločnosti existujú špecializované databázy s vlastným používateľským rozhraním - od návrhu a výpočtov až po logistiku, montáž a prevádzku budovy (obr. 4.4-12). Každý takýto systém spravuje profesionálny tím špecialistov prostredníctvom špeciálneho používateľského rozhrania alebo prostredníctvom databázových dotazov, pričom za sumárom všetkých rozhodnutí o

zadaných hodnotách na konci reťazca stojí správca systému alebo vedúci oddelenia, ktorý zodpovedá za právnu platnosť a kvalitu zadaných údajov svojim protistranám obsluhujúcim ostatné systémy.



Obrázok 4.4-12 Tá istá entita má v rôznych systémoch rovnaký identifikátor, ale rôzne atribúty, ktoré sú dôležité len v danom systéme.

Po usporiadaní zberu štruktúrovaných požiadaviek a údajov na logickej a fyzickej úrovni nám zostáva nastaviť proces automatickej kontroly údajov z rôznych prichádzajúcich dokumentov a rôznych systémov na základe vopred zozbieraných požiadaviek.

### Overenie údajov na stránke a výsledky overenia

Všetky nové údaje vstupujúce do systému - či už ide o dokumenty, tabuľky alebo záznamy v databáze od klienta, architekta, inžiniera, majstra, logistika alebo správcu nehnuteľnosti - musia byť overené na základe predtým formulovaných požiadaviek (obr. 4.4-9). Proces validácie je kritický: akékoľvek chyby v údajoch môžu viest' k nesprávnym výpočtom, oneskoreniu harmonogramu a dokonca k finančným stratám. Na minimalizáciu takýchto rizík by sa mal zorganizovať systematický a opakovateľný, iteračný postup validácie údajov.

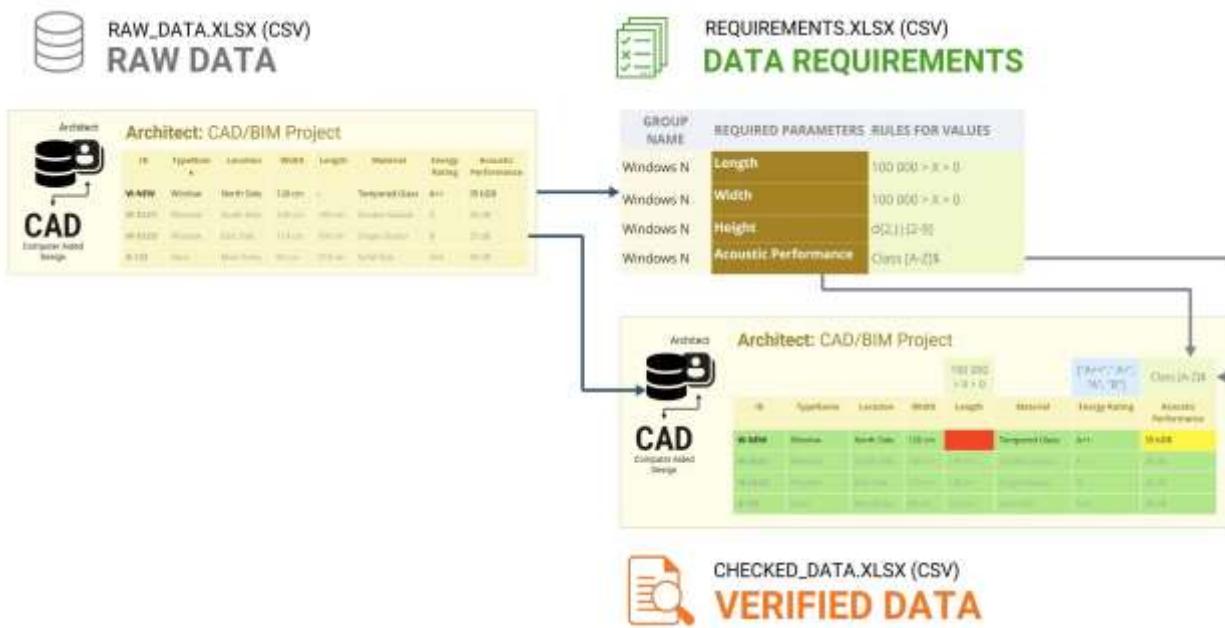
Na overenie nových údajov vstupujúcich do systému - neštruktúrovaných, textových alebo geometrických - je potrebné ich previesť do voľne štruktúrovaného alebo štruktúrovaného formátu. Proces validácie potom musí skontrolovať údaje na základe úplného zoznamu požadovaných atribútov a ich povolených hodnôt.

Konverzia rôznych typov údajov: textu, obrázkov, PDF -dokumentov a zmiešaných údajov CAD (BIM)

do štruktúrovanej podoby bola podrobne opísaná v kapitole "Konverzia údajov do štruktúrovanej podoby".

Príkladom je tabuľka získaná z projektu CAD (BIM) (obr. 4.4-11). Obsahuje pološtruktúrované geometrické údaje a štruktúrované atribútové informácie o entitách projektu (obr. 3.1-14) - napr. prvok z triedy "Windows".

Na vykonanie validácie porovnávame hodnoty atribútov (obr. 4.4-11) s referenčnými hraničnými hodnotami, ktoré boli definované expertmi vo forme požiadavky (obr. 4.4-9). Z výslednej porovnávacej tabuľky (obr. 4.4-13) bude jasné, ktoré hodnoty sú prijateľné a ktoré je potrebné opraviť pred tým, ako sa údaje budú môcť použiť mimo aplikácií CAD (BIM).



Obrázok 4.4-13 V záverečnej validačnej tabuľke sú zvýraznené tie hodnoty atribútov pre novú entitu triedy "Windows", ktorým by ste mali venovať pozornosť.

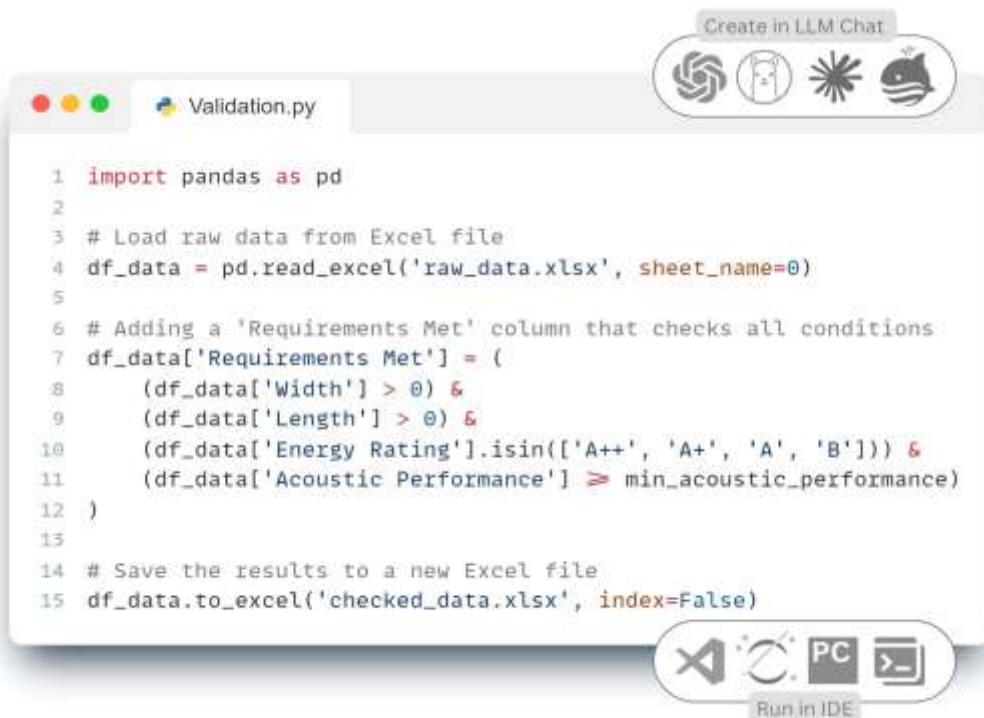
Pri implementácii podobného riešenia pomocou knižnice Pandas, ktorú sme opísali skôr v kapitole "Pandas: Nupostrádateľný nástroj na analýzu údajov", overíme údaje z tabuľkového súboru získaného zo súboru CAD (BIM) (RVT, IFC, DWG, NWS, DGN) (obrázok 4.4-11) pomocou požiadaviek z iného tabuľkového súboru požiadaviek (obrázok 4.4-9).

Na získanie kódu musíme vo výzve pre LLM opísať, že potrebujeme načítať údaje zo súboru **raw\_data.xlsx** (kompletný súbor údajov z databázy CAD (BIM)), **skontrolovať** ich a výsledok uložiť do nového súboru **checked\_data.xlsx** (obr. 4.4-13).

💡 Získajme kód pomocou LLM bez uvedenia knižnice Pandas:

Napíšte kód na kontrolu tabuľky zo súboru raw\_data.xlsx a overte ich pomocou nasledujúcich pravidiel overovania: hodnoty stĺpcov "Width" a "Length" sú väčšie ako nula, "Energy Rating" je zahrnutý v zozname ["A++", "A+", "A", "B"] a "Acoustic Performance" ako premenná, ktorú určíme neskôr - s pridaným posledným stĺpcom overovania, a uložte konečnú tabuľku do nového súboru Excel checked\_data.xlsx ↴

- 💡 V odpovedi LLM sa opíše krátky príklad kódu Pythonu, ktorý sa môže spresniť a rozšíriť pomocou ďalších výziev :



```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

Obrázok 4.4-14 Kód vygenerovaný modelom LLM- kontroluje prevedený návrh CAD (BIM ) podľa požiadaviek na atribúty vo forme hraničných hodnôt.

Kód vygenerovaný jazykovým modelom LLM, možno použiť v akomkoľvek populárnom IDE alebo online nástroji: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo populárne online nástroje Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Vykonanie kódu (obr. 4.4-14) ukáže, že "prvky entity" W-OLD1, W-OLD2, D-122 (a ďalšie prvky) z databázy CAD (BIM) splňajú požiadavky na atribúty: šírka a dĺžka sú väčšie ako nula a trieda energetickej účinnosti je jednou z hodnôt zoznamu "A++", "A", "B", "C" (obr. 4.4-15).

Prvok W-NEW, ktorý potrebujeme a ktorý sme nedávno pridali a ktorý je zodpovedný za novú triedu

prvok "Window" na severnej strane, nie je v súlade s predpismi (atribút "Requirements Met"), pretože jeho dĺžka je nulová (hodnota "0,0" sa podľa nášho pravidla "Width">>0 považuje za neprijateľnú) a neuvádza triedu energetickej účinnosti.



**CHECKED\_DATA.XLSX (CSV)**

**VERIFIED DATA**

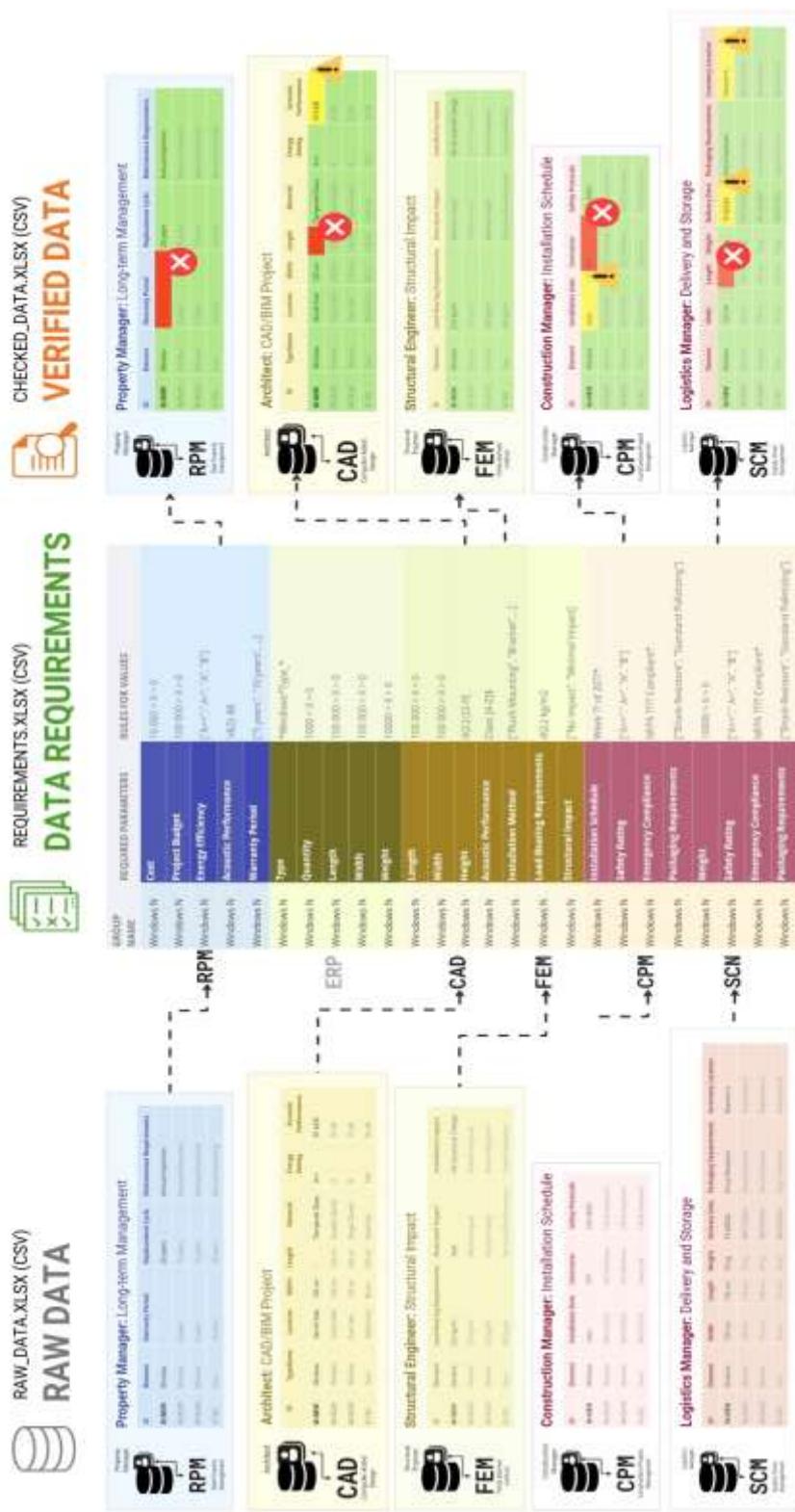
ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met	
0	W-NEW	Window	North Side	120	0.0	Tempered Glass		35	False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++	30	True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B-	25	True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B-	30	True

Obrázok 4.4-15 Kontrola identifikuje entity, ktoré neprešli procesom overovania, a do výsledkov pridá nový atribút s hodnotami "False" (nepravdivý) alebo "True" (pravdivý).

Podobne kontrolujeme konzistenciu všetkých prvkov projektu (entít) a požadovaných atribútov pre každý zo systémov, tabuľiek alebo databáz vo všetkých údajoch, ktoré dostávame od rôznych špecialistov (obrázok 4.4-1) počas procesu pridávania okna do projektu.

V záverečnej tabuľke je vhodné pre vizualizáciu farebne zvýrazniť výsledky kontroly: zelenou farbou sú označené atribúty, ktoré úspešne prešli kontrolou, žltou farbou hodnoty s nekritickými odchýlkami a červenou farbou kritické nezrovnalosti (obr. 4.4-16).

Výsledkom validácie (obrázok 4.4-16) je zoznam dôveryhodných a validovaných prvkov s ich identifikátormi, pri ktorých bolo overené, či spĺňajú požiadavky na atribúty. Validované prvky poskytujú istotu, že tieto prvky spĺňajú stanovené normy a špecifikácie pre všetky systémy zapojené do procesu pridávania prvkov triedy Window alebo akejkoľvek inej triedy (o automatizácii validácie údajov a vytváraní automatizovaného procesu ETL budeme viac hovoriť v kapitole "Automatizácia ETL a validácia údajov").



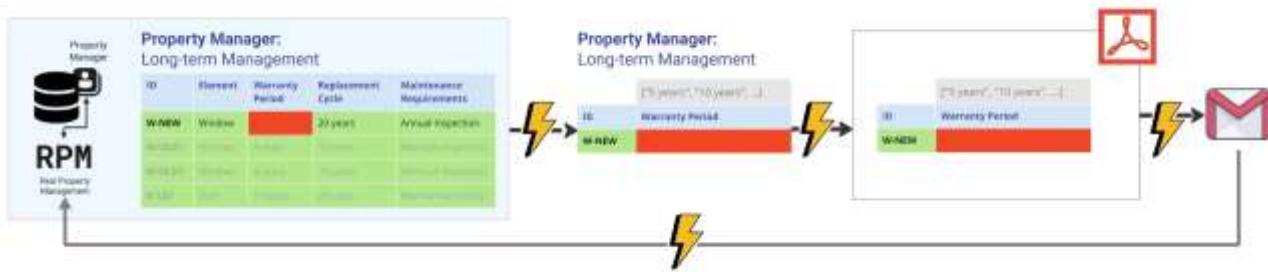
Obr. 4.4-16 Výsledok kontroly vykonanej pre všetky systémy umožňuje určiť, ktoré údaje nespĺňajú požiadavky spoločnosti.

Subjekty, ktoré boli úspešne validované, zvyčajne nevyžadujú veľkú pozornosť. Bez prekážok prechádzajú do ďalších fáz spracovania a integrácie do iných systémov. Na rozdiel od "kvalitných" položiek sú najzaujímavejšie položky, ktoré pri validácii neuspeli. Informácie o takýchto odchýlkach sú veľmi dôležité: mali by sa oznamovať nielen v tabuľkových prehľadoch, ale aj pomocou rôznych vizualizačných nástrojov. Grafické znázornenie výsledkov overovania pomáha rýchlo posúdiť celkový stav kvality údajov, identifikovať problémové oblasti a okamžite priať nápravné alebo korekčné opatrenia.

## Vizualizácia výsledkov overovania

Vizualizácia je základným nástrojom na interpretáciu výsledkov kontroly. Okrem obvyklých súhrnných tabuľiek môže obsahovať aj informačné panely, diagramy a automaticky generované dokumenty vo formáte PDF, ktoré zoskupujú prvky projektu podľa ich stavu kontroly. Farebné kódovanie tu môže zohrávať podpornú úlohu: zelená môže označovať prvky, ktoré boli úspešne overené, žltá môže označovať prvky, ktoré si vyžadujú dodatočnú pozornosť, a červená môže označovať prvky, ktoré majú kritické chyby alebo v nich chýbajú kľúčové údaje.

V našom príklade (obr. 4.4-1) analyzujeme údaje z každého systému krok za krokom: od CAD (BIM) a správy majetku až po logistiku a harmonogramy inštalácie (obr. 4.4-16). Po vykonaní auditu sa pre každého špecialistu automaticky vygenerujú jednotlivé upozornenia alebo dokumenty so správami, napr. vo formáte PDF (obr. 4.4-17). Ak sú údaje správne, špecialista dostane krátku správu: "Ďakujeme za spoluprácu". Ak sa zistia nezrovnalosti, zašle sa podrobnejšia správa so znením: "Tento dokument obsahuje zoznam prvkov, ich identifikátorov, atribútov a hodnôt, ktorých súlad neboli skontrolovaný".



Obrázok 4.4-17 Validácia a automatické generovanie zostáv urýchľuje proces vyhľadávania a pochopenia nedostatkov údajov pre odborníka, ktorý údaje vytvára.

Vďaka automatizovanému procesu validácie - hned' ako sa zistí chyba alebo medzera v údajoch, osobe zodpovednej za vytvorenie alebo spracovanie príslušných entít a ich atribútov sa okamžite odošle oznámenie vo forme správy v chate, e-mailu alebo PDF - dokumentu (obr. 4.4-18) so zoznamom prvkov a opisom atribútov, ktoré neboli validované.



Obr. 4.4-18 Automatické kontrolné správy uľahčujú pochopenie chýb a urýchľujú prácu pri kompletizácii údajov o projekte.

Ak napríklad príde do systému správy majetku dokument (po štruktúrovaní) s nesprávne vyplneným atribútom "Záručná doba", správca majetku dostane upozornenie so zoznamom atribútov, ktoré je potrebné skontrolovať a opraviť.

Podobne akékoľvek nedostatky v harmonograme inštalácie alebo v logistických údajoch budú mať za následok automatické vygenerovanie správy a napríklad odoslanie chatového oznámenia alebo e-mailu s výsledkami kontroly príslušnému špecialistovi.

Okrem PDF -dokumentov a grafov s výsledkami je možné vytvárať informačné panely a interaktívne 3D -modely (obrázok 7.1-6, obrázok 7.2-12), ktoré zvýrazňujú prvky s chýbajúcimi atribútmi, čo umožňuje používateľom vizuálne používať 3D geometriu prvkov na filtrovanie a hodnotenie kvality a úplnosti týchto prvkov v projekte.

Vizualizácia výsledkov kontroly vo forme automaticky generovaných dokumentov, grafov alebo informačných tabuľiek výrazne zjednodušuje interpretáciu údajov a uľahčuje efektívnu komunikáciu medzi účastníkmi projektu.

Proces automatickej kontroly údajov z rôznych systémov a informačných zdrojov možno pripojiť k informovanému rozhodovaniu v každodennom živote. Tak ako firmy v stavebnictve zvažujú množstvo premenných - od spoľahlivosti vstupných údajov až po ich vplyv na čas, náklady a kvalitu realizácie projektu - tak aj človek pri dôležitých rozhodnutiach, napríklad pri výbere miesta na bývanie, zvažuje celý rad faktorov: dopravnú dostupnosť, infraštruktúru, náklady, bezpečnosť, kvalitu života. Všetky tieto úvahy tvoria systém kritérií, ktoré sú základom pre konečné rozhodnutia, z ktorých sa skladá náš život.

## Porovnanie kontrol kvality údajov s potrebami ľudského života

Napriek neustálemu vývoju metód a nástrojov na kontrolu kvality údajov zostáva základný princíp dodržiavania informačných pravidiel nezmenený. Tento princíp je zabudovaný do základov vyspelého systému riadenia, či už v podnikaní alebo v každodennom živote.

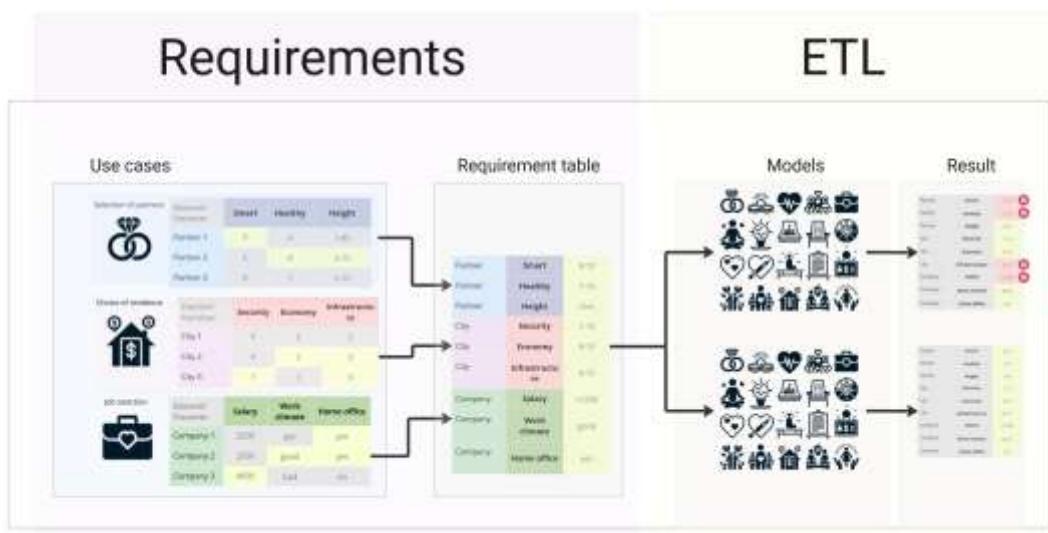
Proces opakovaného overovania údajov sa veľmi podobá rozhodovaciemu procesu, ktorému každý z nás denne čelí. V oboch prípadoch sa spoliehame na skúsenosti, údaje a nové informácie, ktoré sú k dispozícii. A čoraz viac životných a profesionálnych rozhodnutí - od strategických až po každodenné rozhodnutia - sa robí na základe údajov.

Napríklad pri výbere miesta na bývanie alebo životného partnera si v mysli intuitívne vytvárame tabuľku kritérií a vlastností, podľa ktorých porovnávame alternatívy (obr. 4.4-19). Tieto charakteristiky - či už ide o osobné vlastnosti človeka alebo parametre nehnuteľnosti - predstavujú atribúty, ktoré ovplyvňujú konečné rozhodnutie.

Use cases			Requirement table																																																																														
 <p><b>Selection of partners</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Element/Parameter</th> <th style="width: 20%;">Smart</th> <th style="width: 20%;">Healthy</th> <th style="width: 20%;">Height</th> </tr> </thead> <tbody> <tr> <td>Partner 1</td> <td>9</td> <td>6</td> <td>1.65</td> </tr> <tr> <td>Partner 2</td> <td>6</td> <td>8</td> <td>2.10</td> </tr> <tr> <td>Partner 3</td> <td>8</td> <td>7</td> <td>2.10</td> </tr> </tbody> </table>  <p><b>Choice of residence</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Element/Parameter</th> <th style="width: 20%;">Security</th> <th style="width: 20%;">Economy</th> <th style="width: 20%;">Infrastructure</th> </tr> </thead> <tbody> <tr> <td>City 1</td> <td>6</td> <td>5</td> <td>3</td> </tr> <tr> <td>City 2</td> <td>6</td> <td>9</td> <td>9</td> </tr> <tr> <td>City 3</td> <td>7</td> <td>3</td> <td>9</td> </tr> </tbody> </table>  <p><b>Job selection</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Element/Parameter</th> <th style="width: 20%;">Salary</th> <th style="width: 20%;">Work climate</th> <th style="width: 20%;">Home office</th> </tr> </thead> <tbody> <tr> <td>Company 1</td> <td>2200</td> <td>gut</td> <td>yes</td> </tr> <tr> <td>Company 2</td> <td>2500</td> <td>good</td> <td>yes</td> </tr> <tr> <td>Company 3</td> <td>4000</td> <td>bad</td> <td>no</td> </tr> </tbody> </table>	Element/Parameter	Smart	Healthy	Height	Partner 1	9	6	1.65	Partner 2	6	8	2.10	Partner 3	8	7	2.10	Element/Parameter	Security	Economy	Infrastructure	City 1	6	5	3	City 2	6	9	9	City 3	7	3	9	Element/Parameter	Salary	Work climate	Home office	Company 1	2200	gut	yes	Company 2	2500	good	yes	Company 3	4000	bad	no	 <p><b>Requirement table</b></p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">Partner</th> <th style="width: 20%;">Smart</th> <th style="width: 20%;">Healthy</th> <th style="width: 20%;">Height</th> </tr> </thead> <tbody> <tr> <td>Partner</td> <td>8-10</td> <td>7-10</td> <td>&gt;2m.</td> </tr> <tr> <td>Partner</td> <td>7-10</td> <td>7-10</td> <td></td> </tr> <tr> <td>City</td> <td>7-10</td> <td>5-10</td> <td></td> </tr> <tr> <td>City</td> <td>6-10</td> <td>6-10</td> <td></td> </tr> <tr> <td>Company</td> <td>&gt;2300</td> <td>good</td> <td></td> </tr> <tr> <td>Company</td> <td></td> <td>good</td> <td></td> </tr> <tr> <td>Company</td> <td></td> <td>yes</td> <td></td> </tr> </tbody> </table>	Partner	Smart	Healthy	Height	Partner	8-10	7-10	>2m.	Partner	7-10	7-10		City	7-10	5-10		City	6-10	6-10		Company	>2300	good		Company		good		Company		yes	
Element/Parameter	Smart	Healthy	Height																																																																														
Partner 1	9	6	1.65																																																																														
Partner 2	6	8	2.10																																																																														
Partner 3	8	7	2.10																																																																														
Element/Parameter	Security	Economy	Infrastructure																																																																														
City 1	6	5	3																																																																														
City 2	6	9	9																																																																														
City 3	7	3	9																																																																														
Element/Parameter	Salary	Work climate	Home office																																																																														
Company 1	2200	gut	yes																																																																														
Company 2	2500	good	yes																																																																														
Company 3	4000	bad	no																																																																														
Partner	Smart	Healthy	Height																																																																														
Partner	8-10	7-10	>2m.																																																																														
Partner	7-10	7-10																																																																															
City	7-10	5-10																																																																															
City	6-10	6-10																																																																															
Company	>2300	good																																																																															
Company		good																																																																															
Company		yes																																																																															

Obr. 4.4-19 Výber bydliska, zamestnania alebo partnerstva je založený na individuálnych atribútových požiadavkách.

Používanie štruktúrovaných údajov a formalizovaný prístup k opisu požiadaviek (obrázok 4.4-20) prispieva k informovanejším a informovanejším rozhodnutiam v profesionálnom aj osobnom živote.

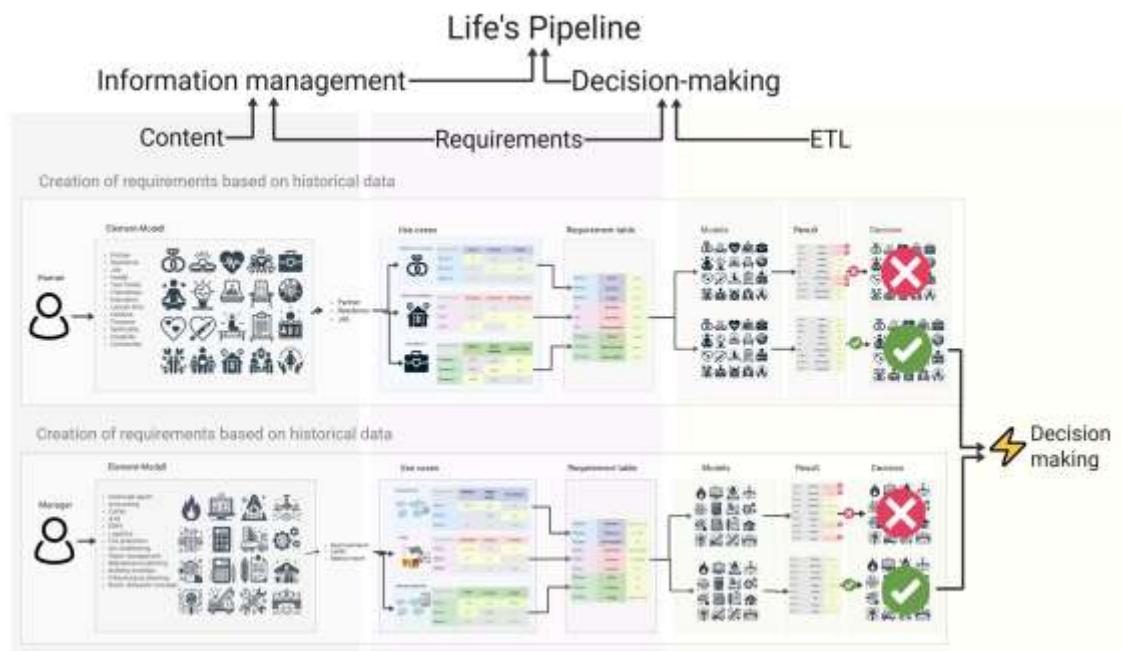


Obr. 4.4-20 Formalizácia požiadaviek pomáha systematizovať vnímanie života a obchodných rozhodnutí.

Prístup k rozhodovaniu založený na údajoch nie je výlučne obchodným nástrojom. Bez problémov sa integruje aj do každodenného života, pričom sa riadi bežnými krokmi spracovania údajov (obr. 4.4-21) podobne ako proces ETL (Extract, Transform, Load), o ktorom sme už hovorili na začiatku tejto časti pri štruktúrovaní údajov a ktorý budeme podrobne skúmať v kontexte automatizácie úloh v siedmej časti knihy:

- **Údaje ako základ (výpis):** V každej oblasti - či už ide o prácu alebo osobný život - zhromažďujeme informácie. V podnikaní to môžu byť správy, čísla, údaje o trhu; v osobnom živote to môžu byť osobné skúsenosti, rady blízkych, spätná väzba, pozorovania.
- **Kritériá hodnotenia (transformácia):** zozbierané informácie sa interpretujú na základe vopred stanovených kritérií. V práci sú to ukazovatele výkonnosti (KPI), rozpočtové obmedzenia a normy; v osobnom živote parametre ako cena, pohodlie, spoľahlivosť, charizma atď.
- **Predpovedanie a analýza rizík (Load):** posledná fáza zahŕňa rozhodovanie na základe analýzy transformovaných údajov a porovnania možných dôsledkov. Je to podobné ako pri obchodných procesoch, kde údaje prechádzajú obchodnou logikou a filtrom rizík.

Rozhodnutia, ktoré robíme - od triviálnych preferencií, ako napríklad čo jest na raňajky, až po dôležité životné udalosti, ako je výber povolania alebo životného partnera - sú vo svojej podstate výsledkom spracovania a vyhodnocovania údajov.



Obr. 4.4-21 Podnikanie a život vo všeobecnosti je sériou rozhodnutí založených na údajoch, pričom kvalita údajov použitých na rozhodovanie je klúčovým faktorom.

Všetko v našom živote je vzájomne prepojené a tak ako sa živé organizmy vrátane človeka riadia prírodnými zákonmi, vyvíjajú sa a prispôsobujú meniacim sa podmienkam, aj ľudské procesy vrátane spôsobu, akým zhromažďujeme a analyzujeme údaje, odrážajú tieto prírodné princípy. Úzky vzťah medzi prírodou a ľudskou činnosťou potvrdzuje nielen našu závislosť od prírody, ale aj našu túžbu uplatňovať zákony zdokonalované miliónmi rokov evolúcie pri vytváraní dátových architektúr, procesov a systémov na rozhodovanie.

Nové technológie, najmä v stavebníctve, sú najlepším príkladom toho, ako sa ľudstvo stále inšpiruje prírodou, aby vytvorilo lepšie, udržateľnejšie a efektívnejšie riešenia.

## Ďalšie kroky: premena údajov na presné výpočty a plány

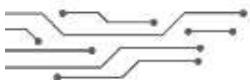
V tejto časti sme sa venovali tomu, ako konvertovať neštruktúrované údaje do štruktúrovaného formátu, vytvárať dátové modely a organizovať procesy kontroly kvality informácií v stavebných projektoch. Správa, štandardizácia a klasifikácia údajov je základný proces, ktorý si vyžaduje systematický prístup a jasné pochopenie obchodných požiadaviek. Techniky a nástroje opísané v tejto časti umožňujú spoľahlivú integráciu medzi rôznymi systémami počas celého životného cyklu objektu.

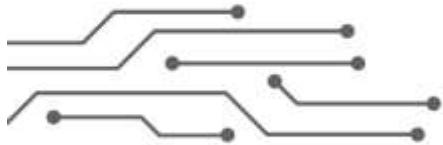
Aby sme zhrnuli túto časť, zdôrazníme hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované prístupy pri vašich každodenných úlohách:

- Začnite systematizáciou požiadaviek
- Vytvorenie registra atribútov a parametrov pre klúčové prvky vašich projektov a procesov

- Zdokumentujte hraničné hodnoty pre každý atribút
- Vizualizovať procesy a vzťahy medzi triedami, systémami a atribútmi pomocou vývojových diagramov (napr. v aplikáciách Miro, Canva, Visio)
- Automatizácia konverzie údajov**
  - Skontrolujte, ktoré z vašich dokumentov, ktoré sa často používajú v procesoch, možno digitalizovať pomocou knižníc OCR a previesť do tabuľkovej formy.
  - Pozrite si nástroje reverzného inžinierstva na extrahovanie údajov z CAD (BIM).
  - Skúste nastaviť automatické načítanie údajov z dokumentov alebo formátov, ktoré často používate pri svojej práci, do tabuľkového formulára.
  - Nastavenie automatických konverzií medzi rôznymi dátovými formátmi
- Vytvorenie databázy znalostí na kategorizáciu**
  - Vyvinúť interný alebo použiť existujúci klasifikátor prvkov v súlade s priemyselnými normami
  - zdokumentovať vzájomné vzťahy medzi rôznymi klasifikačnými systémami
  - Prediskutujte s tímom tému používania jednotného systému identifikácie a jednoznačnej klasifikácie prvkov
  - Začnite budovať proces automatickej validácie údajov - tých, s ktorými pracujete v rámci tímu, aj tých, ktoré sa odovzdávajú do externých systémov.

Pomocou týchto prístupov môžete výrazne zlepšiť kvalitu svojich údajov a zjednodušiť ich následné spracovanie a transformáciu. V nasledujúcich častiach tejto knihy sa budeme venovať tomu, ako použiť už štruktúrované a pripravené údaje na automatizované výpočty, odhadovanie nákladov, plánovanie a riadenie stavebných projektov.





## V ČASŤ VÝPOČTY NÁKLADOV A ČASU: ZAČLENENIE ÚDAJOV DO STAVEBNÝCH PROCESOV

Piata časť je venovaná praktickým aspektom využívania údajov na optimalizáciu kalkulácie nákladov a plánovania stavebných projektov. Podrobne sa analyzuje metóda odhadu nákladov založená na zdrojoch a automatizácia procesov odhadu. Uvažuje sa o metódach automatizovaného získavania Quantity Take-Off) z CAD (BIM)-modelov a ich integrácie s kalkulačnými systémami. Skúmajú sa technológie 4D a 5D modelovania pre časové plánovanie a riadenie stavebných nákladov s konkrétnymi príkladmi ich použitia. Uvádzia sa analýza rozšírených informačných vrstiev 6D -8D, ktoré poskytujú integrovaný prístup k posudzovaniu udržateľnosti, prevádzky a bezpečnosti objektov nehnuteľností. Podrobne sa rozoberajú metódy výpočtu uhlíkovej stopy a ESG -indikátorov stavebných projektov v kontexte moderných environmentálnych požiadaviek a noriem. Kriticky sa hodnotia možnosti a obmedzenia tradičných systémov ERP a PMIS pri riadení stavebných procesov s analýzou ich vplyvu na transparentnosť tvorby cien. Predpokladajú sa perspektívy prechodu od uzavretých riešení k otvoreným štandardom a flexibilným nástrojom na analýzu údajov, ktoré dokážu zabezpečiť vyššiu efektívnosť stavebných procesov.

# KAPITOLA 5.1.

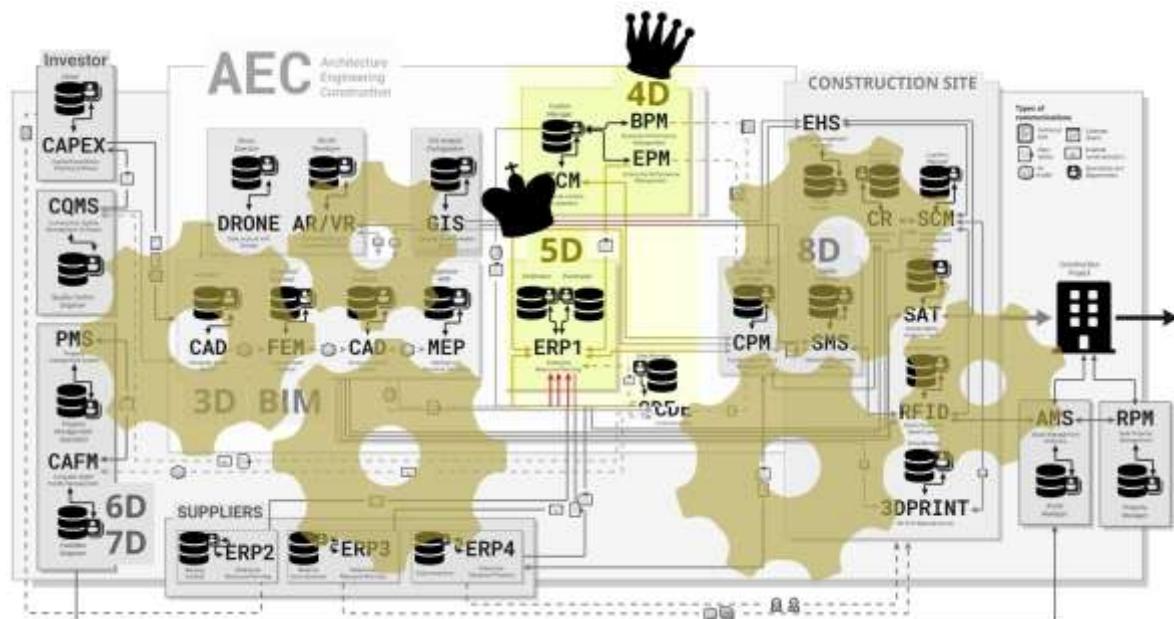
## VÝPOČTY A ODHADY NÁKLADOV NA STAVEBNÉ PROJEKTY

## Základy stavebníctva: odhad množstva, nákladov a času

Spomedzi mnohých podnikových procesov, ktoré rozhodujú o udržateľnosti podniku v stavebníctve, majú mimoriadny význam - tak ako pred tisícami rokov - procesy presného odhadu počtu prvkov, nákladov na projekt a času dodania (obr. 5.1-1).

Vývoj písma bol výsledkom viacerých faktorov vrátane potreby zaznamenávať hospodárske transakcie, obchod a hospodárenie so zdrojmi v raných spoločnostiach. Prvé právne významné dokumenty, hlinené tabuľky s výpočtom materiálových nákladov a mzdových nákladov, sa používali v súvislosti s obchodom a stavebníctvom. Tieto tabuľky zaznamenávali záväzky strán pri výstavbe stavieb a uchovávali sa ako dôkaz dohôd a peňažných a obchodných vzťahov.

Po tisícročia zostával prístup k odhadom zväčša nezmenený: výpočty sa vykonávali ručne a spoliehali sa na skúsenosti a intuíciu odhadujúceho inžiniera. S príchodom modulárnych systémov ERP a nástrojov CAD sa však tradičný prístup k odhadovaniu množstva, nákladov a času začal rýchlo meniť. Dnešné digitálne technológie umožňujú plne automatizovať kľúčové výpočty času a nákladov, čo umožňuje väčšiu presnosť, rýchlosť a transparentnosť pri plánovaní zdrojov stavebných projektov.



Obr. 5.1-1 Z množstva rôznych systémov sú v podnikaní najdôležitejšie nástroje zodpovedné za meranie objemu, nákladov a času.

Stavebné spoločnosti sa zameriavajú na presné údaje o čase a nákladoch. Tie zase závisia od množstva použitého materiálu a práce a ich prehľadnosť ovplyvňuje ziskosť. Zložitosť kalkulačných procesov a ich netransparentnosť však často vedú k vyšším nákladom na projekt, nedodržaniu termínov a dokonca k bankrotu.

Podľa správy KPMG "Známe problémy - nové prístupy" (2023) je len 50% stavebných projektov dokončených načas a 87% spoločností uvádzajú zvýšenú kontrolu nad ekonomikou investičných projektov. Hlavné problémy súvisia s nedostatkom kvalifikovaného personálu a s ťažkosťami pri predpovedaní rizík [2].

Historické údaje o nákladoch a časoch procesov sa zhromažďujú počas výstavby minulých projektov počas celej existencie stavebnej spoločnosti a zadávajú sa do databáz rôznych systémov (ERP, PMIS BPM, EPM atď.).

Kvalitné historické údaje o nákladoch sú pre stavebnú organizáciu veľkou konkurenčnou výhodou, ktorá priamo ovplyvňuje jej schopnosť prežiť.

Oddelenia odhadu a kalkulácie nákladov v stavebných a inžinierskych spoločnostiach sú vytvorené na zhromažďovanie, uchovávanie a aktualizáciu historických údajov o výpočtoch projektov. Ich hlavnou úlohou je zhromažďovať a systematizovať skúsenosti spoločnosti, čo pomáha časom zlepšiť presnosť odhadu rozsahu, termínov a nákladov nových projektov. Tento prístup pomáha minimalizovať chyby v budúcich výpočtoch na základe praxe a výsledkov už realizovaných projektov.

## Metódy výpočtu odhadovaných nákladov na projekty

Odborníci na kalkuláciu nákladov používajú rôzne metódy odhadu, pričom každá z nich sa zameriava na konkrétny typ údajov, dostupnosť informácií a úroveň podrobnosti projektu. Medzi najbežnejšie patria:

- **Metóda založená na zdrojoch:** odhadovanie odhadovaných nákladov na projekt na základe podrobnej analýzy všetkých potrebných zdrojov, ako sú materiály, vybavenie a práca. Táto metóda si vyžaduje podrobný zoznam všetkých úloh a zdrojov potrebných na vykonanie každej úlohy, po ktorom nasleduje výpočet ich nákladov. Táto metóda je veľmi presná a je široko používaná pri odhadovaní nákladov.
- **Parametrická metóda:** používa štatistické modely na odhad nákladov na základe parametrov projektu. Môže ísť o analýzu nákladov na mernú jednotku, ako je plocha budovy alebo rozsah prác, a prispôsobenie týchto hodnôt konkrétnym podmienkam projektu. Táto metóda je účinná najmä v počiatočných fázach, keď ešte nie sú k dispozícii podrobnej informácie.
- **Jednotková metóda (metóda jednotkových nákladov):** vypočíta odhadované náklady projektu na základe nákladov na mernú jednotku (napr. na meter štvorcový alebo meter kubický). Poskytuje rýchly a jednoduchý spôsob porovnania a analýzy nákladov na rôzne projekty alebo časti projektov.
- **Odborný posudok (delfská metóda):** založený na názoroch odborníkov, ktorí na odhad hodnoty projektu využívajú svoje skúsenosti a znalosti. Tento prístup je užitočný, ak nie sú k dispozícii presné východiskové údaje alebo ak je projekt jedinečný.

Treba poznamenať, že parametrickú metódu a expertné posudky možno prispôsobiť modelom strojového učenia. To umožňuje automatické generovanie prognóz nákladov a harmonogramu projektu na základe tréningových vzoriek. Príklady použitia takýchto modelov sú podrobnejšie opísané v kapitole "Príklad použitia strojového učenia na vyhľadávanie nákladov a harmonogramov projektu" (obr. 9.3-5).

Napriek tomu metóda založená na zdrojoch zostáva vo svetovej praxi najpopulárnejšou a najpoužívanejšou. Poskytuje nielen presné posúdenie odhadovaných nákladov, ale umožňuje aj výpočet trvania jednotlivých procesov na stavbe a celého projektu ako celku (podrobnejšie v kapitole "Harmonogramy výstavby a 4D -údaje o projekte").

## **Metóda založená na zdrojoch kalkulácia nákladov a odhad v stavebníctve**

Kalkulácia nákladov na základe zdrojov je metóda manažérskeho účtovníctva, pri ktorej sa náklady na projekt zakladajú na priamom zúčtovaní všetkých použitých zdrojov. V stavebníctve tento prístup zahŕňa podrobnú analýzu a hodnotenie všetkých materiálových, pracovných a technických zdrojov potrebných na vykonanie prác.

Metóda založená na zdrojoch, poskytuje vysoký stupeň transparentnosti a presnosti pri plánovaní rozpočtu, pretože sa zameriava na skutočné ceny zdrojov v čase odhadu. To je obzvlášť dôležité v nestabilnom ekonomickej prostredí, kde výkyvy cien môžu výrazne ovplyvniť celkové náklady projektu.

V nasledujúcich kapitolách sa budeme podrobne zaoberať procesom kalkulácie nákladov na základe zdrojov. Aby sme lepšie pochopili jeho princípy v stavebníctve, urobíme analógiu s kalkuláciou nákladov na večeru v reštaurácii. Manažér reštaurácie pri plánovaní večera zostaví zoznam potrebných produktov, zohľadní čas prípravy jednotlivých jedál a potom náklady vynásobí počtom hostí. V stavebníctve je postup podobný: pre každú kategóriu prvkov projektu (objektov) sa vytvoria položkové odhady Recepty, a celkové náklady projektu sa určia súčtom všetkých nákladov v celkovom vyúčtovaní - konečnom odhade podľa kategórií.

Klúčovou a počiatočnou fázou prístupu založeného na zdrojoch je vytvorenie počiatočnej databázy spoločnosti. V prvej fáze kalkulácie nákladov sa zostaví štruktúrovaný zoznam všetkých položiek, materiálov, druhov prác a zdrojov, ktoré má spoločnosť v rámci svojich stavebných projektov k dispozícii - od klinca v sklede až po opis ľudí cez ich kvalifikáciu a hodinovú sadzbu. Tieto informácie sa systematizujú do jednotnej "Databázy stavebných zdrojov a materiálov" - tabuľkového registra, ktorý obsahuje údaje o názvoch, charakteristikách, merných jednotkách a aktuálnych cenách. Práve táto databáza sa stáva hlavným a primárnym zdrojom informácií pre všetky následné výpočty zdrojov - nákladov aj termínov prác.

## **Databáza stavebných zdrojov: katalóg stavebných materiálov a prác**

Databáza alebo tabuľka stavebných zdrojov a materiálov - obsahuje podrobne informácie o každom

prvku, ktorý sa môže použiť v stavebnom projekte - výrobok, položka, materiál alebo služba, vrátane jeho názvu, opisu, mernej jednotky a jednotkovej ceny, zaznamenané v štruktúrovanej forme. V tejto tabuľke nájdete všetko od rôznych druhov palív a materiálov používaných v projektoch až po podrobne zoznamy špecialistov v podobe rôznych kategórií s popisom hodinových sadzieb (obr. 5.1-2).

## Database of resources

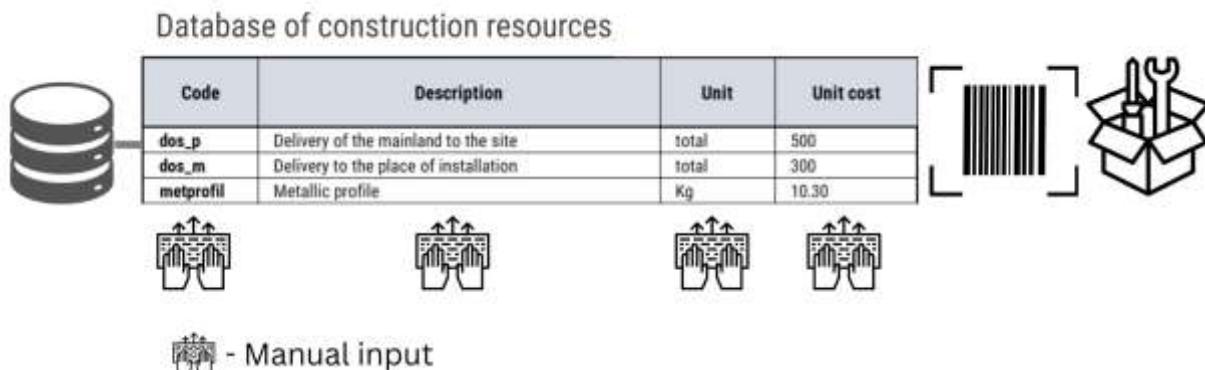
 <p><b>1st grade potatoes</b></p> <p><b>1 kg   \$2,99</b></p>	 <p><b>Sand lime bricks</b></p> <p><b>1 pcs   \$1</b></p>
 <p><b>Black Angus marble beef</b></p> <p><b>1 kg   \$26,99</b></p>	 <p><b>JCB 3CX backhoe loader</b></p> <p><b>1 h   \$150</b></p>
 <p><b>Broccoli</b></p> <p><b>1 pcs   \$1,99</b></p>	 <p><b>Laborer of the 1st category</b></p> <p><b>1 h   \$30</b></p>

Obrázok 5.1-2 Tabuľka zdrojov je zoznam zložiek popisujúci materiál a službu s jednotkovou cenou.

"Databáza zdrojov" je podobná katalógu produktov v internetovom obchode, kde má každá položka podrobny opis svojich vlastností. To uľahčuje odhadcom nákladov výber správnych zdrojov (podobne ako výber produktov pri pridávaní do nákupného košíka) potrebných na výpočet konkrétnych stavebných postupov vo forme kalkulácií (konečné poradie v internetovom obchode).

Databázu zdrojov si možno predstaviť aj ako zoznam všetkých zložiek v kuchárskej knihe reštaurácie. Každý stavebný materiál, zariadenie a služba sú podobné prísadám používaným v receptoch. "Databáza zdrojov" je podrobny zoznam všetkých ingrediencií - stavebných materiálov a služieb vrátane ich jednotkových nákladov: kus, meter, hodina, liter atď.

Nové prvky entít možno do tabuľky "Databázy stavebných zdrojov" pridať dvoma spôsobmi - ručne (obr. 5.1-3) alebo automaticky integráciou so systémami riadenia zásob spoločnosti alebo databázami dodávateľov.



Obr. 5.1-3 Databáza zdrojov sa vypíňa ručne alebo automaticky preberá údaje z iných databáz.

Typická stredne veľká stavebná spoločnosť využíva databázu obsahujúcu tisíce a niekedy až desaťtisíce položiek s podrobnejším popisom, ktoré možno použiť v stavebných projektoch. Tieto údaje sa potom automaticky využívajú v zmluvách a projektovej dokumentácii na presný opis skladby práv a procesov

Aby bolo možné držať krok s meniacimi sa podmienkami na trhu, ako je napríklad inflácia, atribút "jednotkových nákladov" pre každý produkt (tovar alebo službu) v databáze zdrojov (obrázok 5.1-3) sa pravidelne aktualizuje manuálne alebo automatickým preberaním aktuálnych cien z iných systémov alebo online platforem.

Aktualizácia jednotkových nákladov na zdroj sa môže vykonávať mesačne, štvrtročne alebo ročne - v závislosti od povahy zdroja, inflácie a vonkajšej hospodárskej situácie. Takéto aktualizácie sú potrebné na zachovanie presnosti výpočtov a odhadov, keďže ide o základné prvky, ktoré slúžia ako východisko pre prácu odhadcov nákladov. Aktuálne údaje sa používajú na tvorbu odhadov, rozpočtov a harmonogramov, ktoré odrážajú skutočné podmienky na trhu a znižujú riziko chýb v následných výpočtoch projektu.

## Zostavenie výpočtov a výčíslenie nákladov na práce na základe zdrojov

Po naplnení "Databázy stavebných zdrojov" (obr. 5.1-3) minimálnymi jednotkami môžete začať vytvárať výpočty, ktoré sa vypočítajú pre každý proces alebo prácu na stavenisku pre určité merné jednotky: napríklad pre jeden kubický meter betónu, jeden štvorcový meter sadrokartónovej steny, pre jeden meter obrubníka alebo pre montáž okna.

Napríklad na postavenie tehlovej steny s plochou 1 m<sup>2</sup> (obrázok 5.1-4) je na základe skúseností z predchádzajúcich projektov potrebných približne 65 tehál (subjekt "Silikátová tehla") s cenou 1 USD za kus (atribút "Náklady na kus"), spolu 65 USD. Podľa mojich skúseností je tiež potrebné použiť stavebné zariadenie (entita "JCB 3CX Loader") na 10 minút, ktoré umiestni tehly v blízkosti pracovnej oblasti. Keďže prenájom zariadenia stojí 150 USD za hodinu, 6 minút používania by stalo približne 15 USD. Okrem toho bude potrebný dodávateľ na kladenie tehál na 2 hodiny s hodinovou sadzbou 30 USD a celkovou sumou 60 USD.

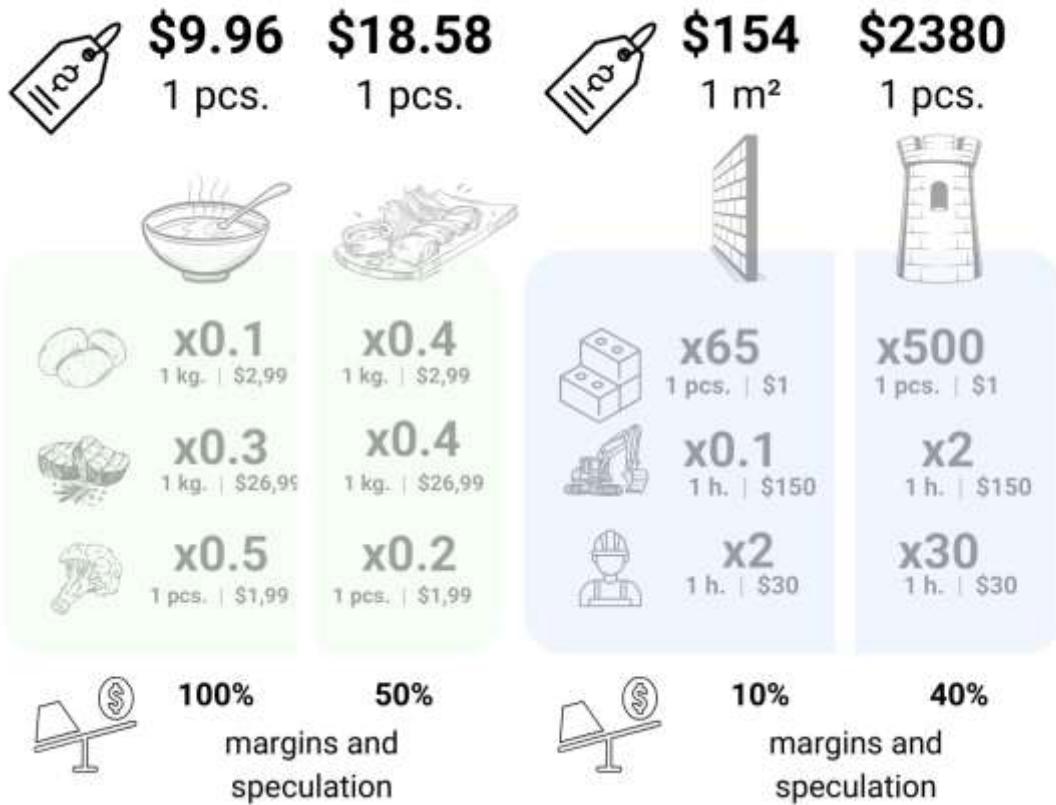
<b>x0.1</b> 1 kg   \$2,99	<b>x0.4</b> 1 kg.   \$2,99	<b>x65</b> 1 pcs   \$1	<b>x500</b> 1 pcs   \$1
<b>x0.3</b> 1 kg   \$26,99	<b>x0.4</b> 1 kg.   \$26,99	<b>x0.1</b> 1 h   \$150	<b>x2</b> 1 h   \$150
<b>x0.5</b> 1 pcs   \$1,99	<b>x0.2</b> 1 pcs.   \$1,99	<b>x2</b> 1 h   \$30	<b>x30</b> 1 h   \$30
+ margins and speculation	+ margins and speculation	+ margins and speculation	+ margins and speculation

Obr. 5.1-4 Kalkulácia nákladov obsahuje podrobný zoznam stavebných materiálov a služieb potrebných na vykonanie prác a procesov.

Zloženie výpočtov (tzv. "receptúr") sa tvorí na základe historických skúseností, ktoré spoločnosť získala pri vykonávaní veľkého objemu podobných prác. Tieto praktické skúsenosti sa zvyčajne zhromažďujú prostredníctvom spätej väzby zo stavby. Predovšetkým majster zbiera informácie priamo na stavbe, zaznamenáva skutočné náklady na prácu, spotrebu materiálu a nuansy technologických operácií. V spolupráci s oddelením odhadu sa potom tieto informácie iteratívne spresňujú: spresňujú sa opisy procesov, upravuje sa skladba zdrojov a aktualizujú sa kalkulácie nákladov tak, aby odrážali skutočné údaje z posledných projektov.

Podobne ako recept opisuje prísady a množstvá potrebné na prípravu pokrmu, odhad nákladov poskytuje podrobný zoznam všetkých stavebných materiálov, zdrojov a služieb potrebných na dokončenie konkrétnej práce alebo procesu.

Pravidelné vykonávaná práca umožňuje pracovníkom, majstrom a odhadcom orientovať sa v potrebnom množstve zdrojov: materiálu, pohonného hmôtu, pracovného času a ďalších parametrov potrebných na vykonanie jednotky práce (obr. 5.1-5). Tieto údaje sa zadávajú do odhadových systémov vo forme tabuľiek, kde je každá úloha a operácia opísaná prostredníctvom minimálnych prvkov zdrojovej základne (s neustále aktualizovanými cenami), čo zabezpečuje presnosť výpočtov.



Obr. 5.1-5 Jednotkové sadzby sa zhromažďujú pre každú úlohu, kde sa atribút objemu entity vynásobí jej množstvom a pripočíta sa percento zisku.

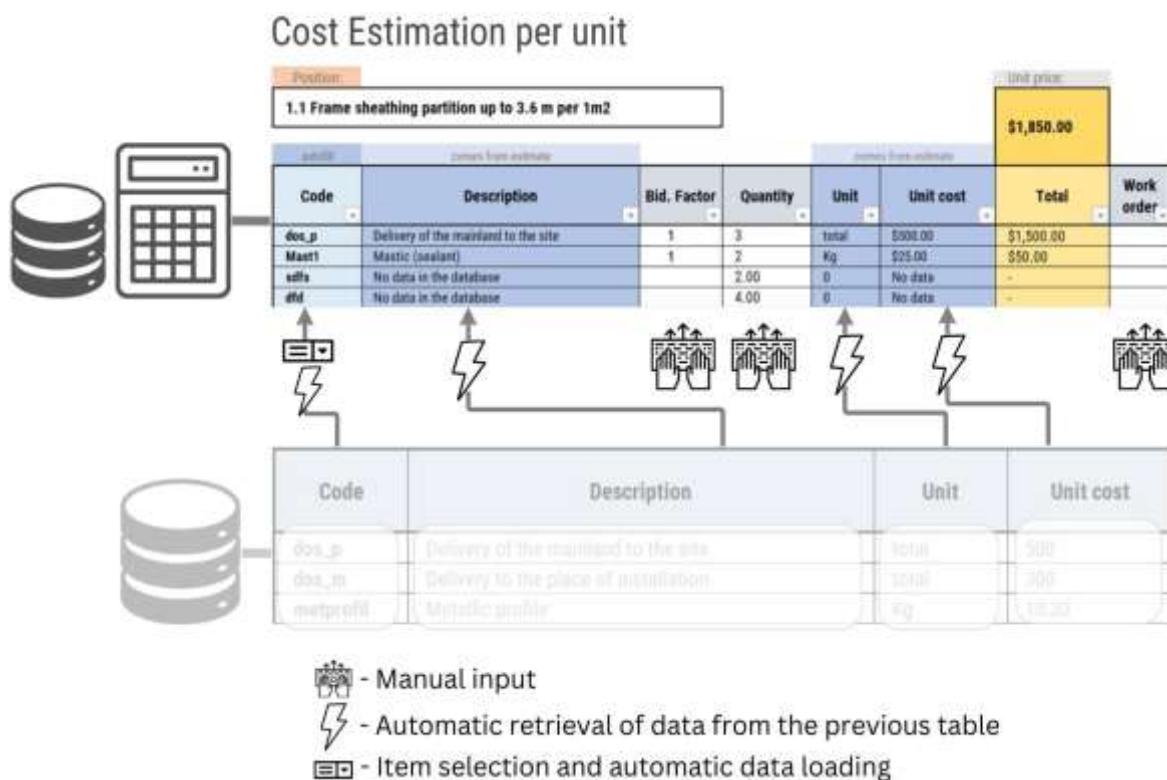
Na získanie celkových nákladov každého procesu alebo činnosti (objektu kalkulácie nákladov) sa atribút nákladov vynásobí ich počtom a koeficientmi. Koeficienty môžu zohľadňovať rôzne faktory, napríklad zložitosť práce, regionálne charakteristiky, mieru inflácie, potenciálne riziká (očakávané percento režie) alebo špekulácie (faktor dodatočného zisku).

Odhadca ako analytik premieňa skúsenosti a odporúčania majstra na štandardizované odhady, ktoré opisujú stavebné procesy prostredníctvom zdrojových entít v tabuľkovej forme. Úlohou odhadcu je v podstate zhromažďovať a štruktúrovať prostredníctvom parametrov a koeficientov informácie prichádzajúce zo staveniska.

Konečné náklady na jednotku práce (napr. štvorcový alebo kubický meter alebo jedna montáž jednej jednotky) teda zahŕňajú nielen priame náklady na materiál a prácu, ale aj podnikové prirážky, režijné náklady, poistenie a ďalšie faktory (obrázok 5.1-6)

Zároveň sa už nemusíme staráť o skutočné ceny pri výpočtoch (receptov), pretože skutočné ceny sa vždy odrážajú v "základni zdrojov" (tabuľke zložiek). Na úrovni výpočtov sa do tabuľky automaticky načítajú údaje zo surovinovej základne (napr. podľa kódu položky alebo jej jedinečného identifikátora), do ktorej sa načíta popis a skutočná cena za jednotku, ktorá sa zase môže automaticky načítať z online

platforem alebo online obchodu so stavebnými materiálmi. Odhadca na úrovni kalkulácií práč musí len popísť prácu alebo proces prostredníctvom atribútu "množstvo zdrojov" a ďalších faktorov.



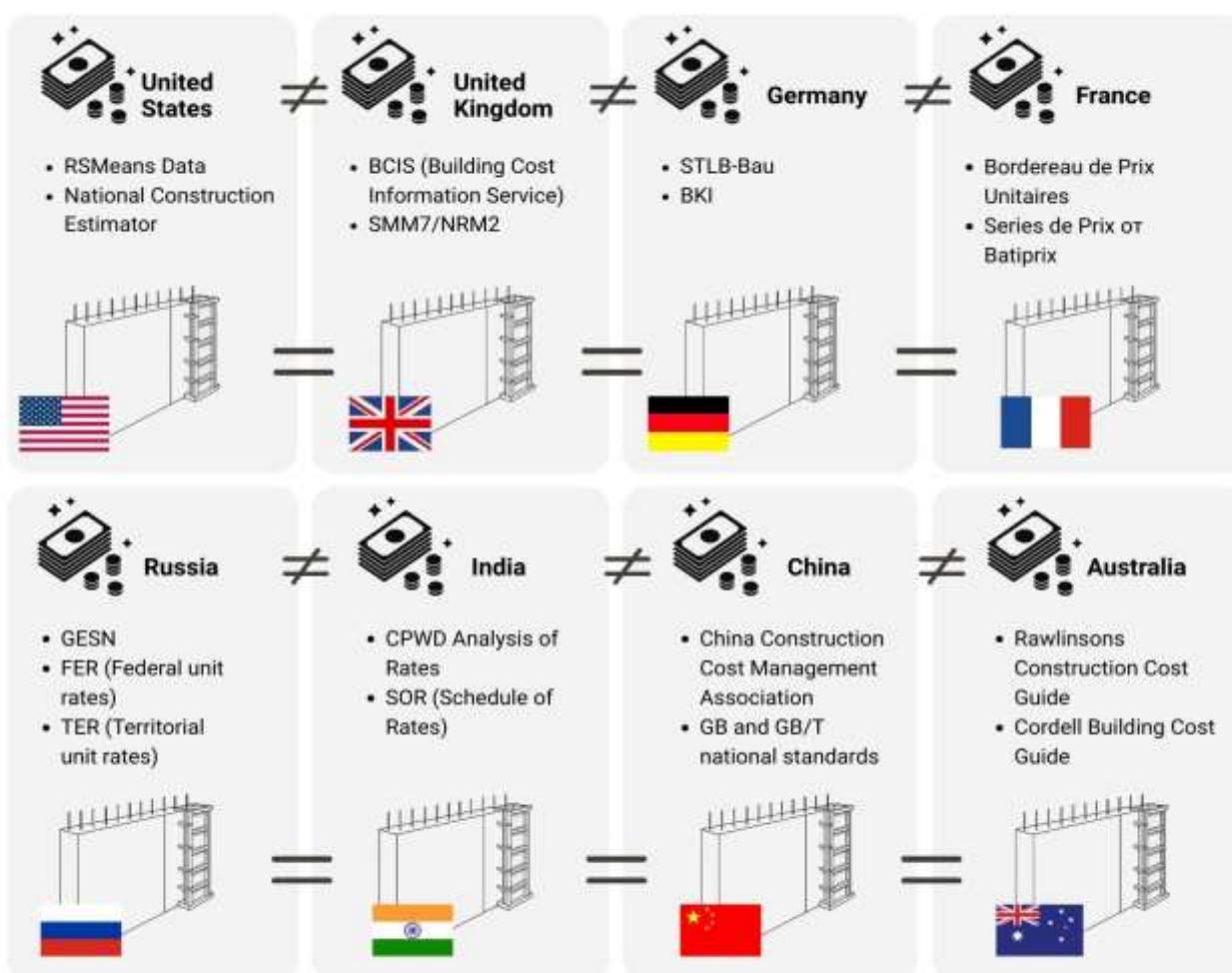
Obr. 5.1-6 Vo fáze výpočtu jednotkových nákladov práce sa vypĺňajú len atribúty množstva potrebných zdrojov, všetko ostatné sa automaticky načíta z databázy zdrojov.

Vytvorené kalkulácie sú uložené vo forme vzorových tabuľiek typických projektov, ktoré sú priamo prepojené s databázou stavebných zdrojov a materiálov. Tieto šablóny predstavujú štandardizované recepty pre opakujúce sa typy práč pre budúce projekty, čím sa zabezpečuje jednotnosť výpočtov v celej spoločnosti.

Ked' sa v databáze zmenia náklady na akýkol'vek zdroj (obr. 5.1-3) - či už manuálne alebo automaticky prostredníctvom preberania aktuálnych trhových cien (napr. v podmienkach inflácie) - aktualizácie sa okamžite premietnu do všetkých prepojených kalkulácií nákladov (obr. 5.1-6). To znamená, že je potrebné meniť len základnú zdrojov, zatiaľ čo šablóny a odhady nákladov zostávajú v priebehu času nezmenené. Tento prístup zabezpečuje stabilitu a reprodukovateľnosť výpočtov pri akýchkoľvek cenových výkyvoch, ktoré sa zohľadňujú len v pomerne jednoduchej tabuľke zdrojov (obr. 5.1-3).

Pre každý nový projekt sa vytvorí kópia štandardnej šablóny kalkulácie nákladov, ktorá umožňuje vykonávať zmeny a upravovať činnosti podľa konkrétnych požiadaviek bez toho, aby sa zmenila pôvodná šablóna prijatá spoločnosťou. Tento prístup poskytuje flexibilitu pri úprave kalkulácií: môžete zohľadniť špecifika staveniska, želania zákazníka, zaviesť koeficienty rizika alebo ziskovosti (špekulácie) - to všetko bez porušenia firemných štandardov. To pomáha spoločnosti nájsť rovnováhu medzi maximalizáciou zisku, spokojnosťou zákazníka a zachovaním konkurencieschopnosti.

V niektorých krajinách sú takéto vzory kalkulácií nákladov, ktoré sa zhromažďujú desaťročia, štandardizované na národnej úrovni a stávajú sa súčasťou noriem národného systému kalkulácií nákladov na výstavbu (obrázok 5.1-7).



Obr. 5.1-7 Rôzne krajin sveta majú vlastné pravidlá kalkulácie nákladov s vlastnými (predpisanými) kompendiami a normami pre stavebné práce na kalkuláciu nákladov toho istého prvku.

Takéto štandardizované zdrojové základne odhadov (obrázok 5.1-7) sú povinné pre všetkých

účastníkov trhunajmä pre projekty financované z verejných zdrojov. Takáto štandardizácia zabezpečuje transparentnosť, porovnatelnosť a spravodlivosť pri tvorbe cien a zmluvných záväzkov pre klienta.

## Konečné vyčíslenie nákladov na projekt: od odhadov k rozpočtom

Štátne a odvetvové normy pre odhadovanie zohrávajú v stavebnej praxi v rôznych krajinách rôznu úlohu. Zatiaľ čo niektoré krajinu vyžadujú prísne dodržiavanie jednej normy, väčšina vyspelých ekonomík uplatňuje flexibilnejší prístup. V trhových ekonomikách slúžia štátne stavebné normy zvyčajne len ako východisko. Stavebné spoločnosti tieto normy prispôsobujú svojim modelom fungovania alebo ich úplne revidujú a dopĺňajú vlastnými prispôsobenými faktormi. Tieto úpravy odrážajú firemné skúsenosti, efektívnosť riadenia zdrojov a často aj faktory, v ktorých sa môžu zohľadniť napríklad špekulatívne zisky spoločnosti.

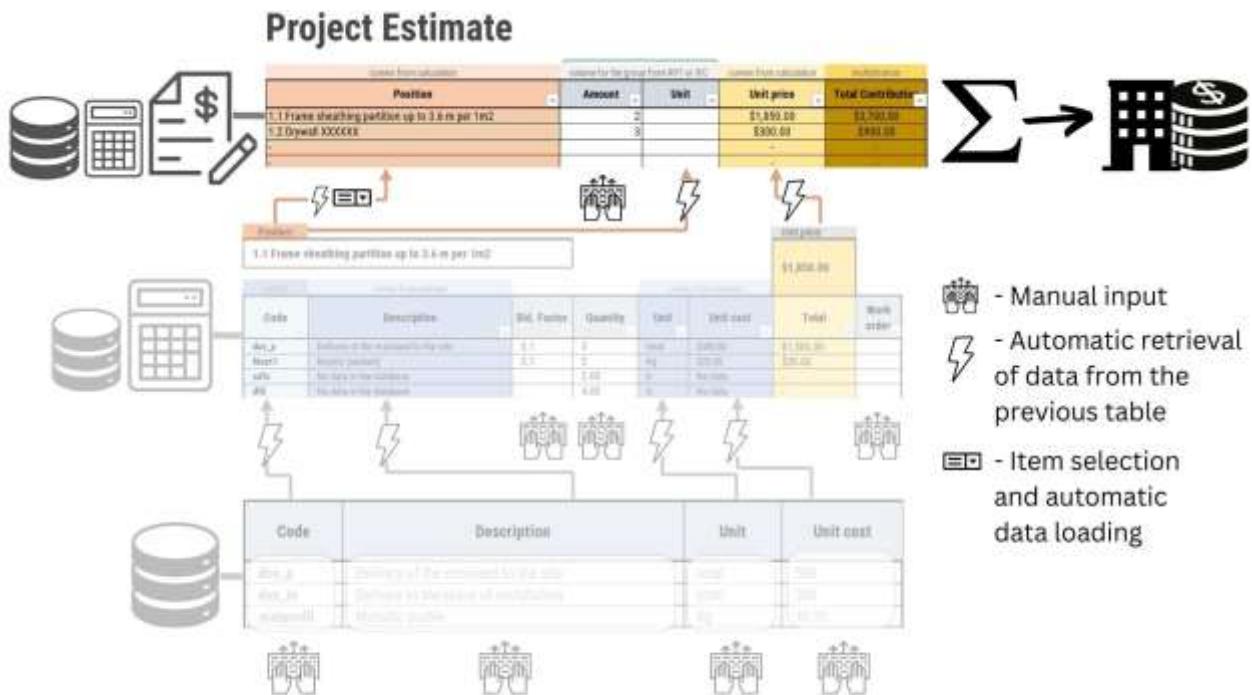
V dôsledku toho môže úroveň konkurencie, dopyt na trhu, cieľové marže a dokonca aj vzťahy s konkrétnymi zákazníkmi viesť k výrazným odchýlkom od štandardizovaných nariem. Tento postup poskytuje flexibilitu trhu, ale zároveň sťahuje transparentné porovnanie ponúk rôznych dodávateľov, čím sa do stavebníctva v tejto fáze kalkulačného procesu zavádzza prvok špekulatívnej tvorby cien.

Po príprave kalkulačných vzorov pre jednotlivé činnosti a procesy - alebo častejšie jednoducho skopírovaných zo štandardných štátnych odhadov (obr. 5.1-7) s pridanými koeficientmi, ktoré odrážajú "specifíká" konkrétnej spoločnosti - je posledným krokom vynásobenie nákladov každej položky príslušným atribútom rozsahu prác alebo procesov v novom projekte.

Pri výpočte celkových nákladov nového stavebného projektu je kľúčovým krokom sumarizácia nákladov všetkých položiek kalkulácie nákladov, vynásobená objemom týchto položiek prác v projekte.

Na vytvorenie celkových nákladov projektu v našom zjednodušenom príklade začneme výpočtom nákladov na vybudovanie jedného štvorcového metra steny a náklady na jej výpočet (napr. práca "1m<sup>2</sup> štandardná inštalácia stenových prvkov") vynásobíme celkovým počtom štvorcových metrov stien v projekte (napr. atribút "Plocha" alebo "Množstvo" (obr. 5.1-8) entity typu "Stenové prvky" z CAD projektu alebo výpočty majstra).

Náklady na všetky prvky projektu vypočítame rovnakým spôsobom (obr. 5.1-8): vezmeme náklady na jednotku práce a vynásobíme ich objemom konkrétneho prvku alebo skupiny prvkov v projekte. Odhadca musí len zadať počet týchto prvkov, činností alebo procesov v projekte na vo forme objemu alebo množstva. To umožňuje automaticky vygenerovať kompletnejší stavebný odhad.



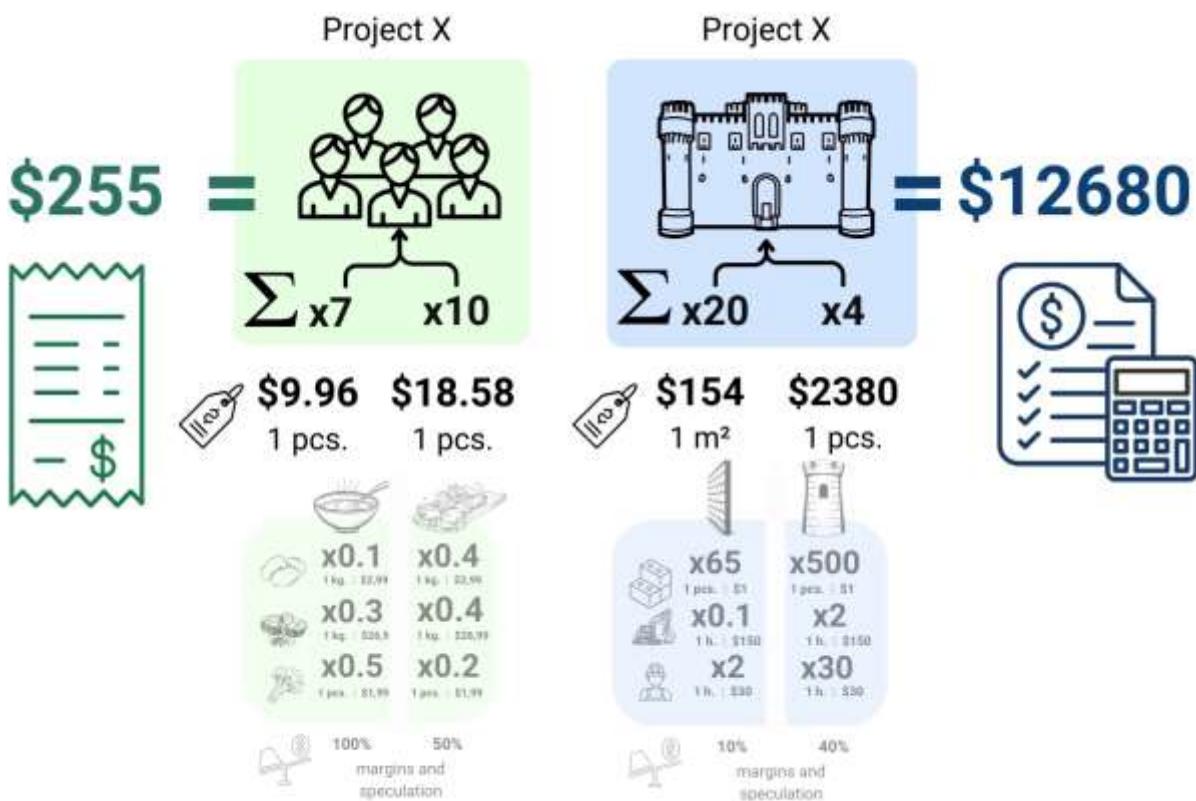
Obr. 5.1-8 Vo fáze tvorby odhadu zadávame len rozsah prác.

Podobne ako v prípade kalkulácií, aj na tejto úrovni nahrávame automaticky pripravené kalkulované položky (zo šablóny kalkulácií alebo nové skopírované zo šablóny a upravené), ktoré so sebou automaticky prinášajú aktuálne náklady na jednotku práce (ktoré sa automaticky aktualizujú z databázy zdrojov (obr. 5.1-8 spodná tabuľka)). V súlade s tým sa v prípade akejkoľvek zmeny údajov v databáze zdrojov alebo v tabuľkách kalkulácie nákladov - údaje v kalkulácii automaticky aktualizujú pre aktuálny deň bez toho, aby bolo potrebné meniť kalkuláciu nákladov alebo samotnú kalkuláciu.

V reštauračnom kontexte sa konečné náklady na podujatie vypočítavajú podobným spôsobom a rovnajú sa konečným nákladom na celú večeru, pričom náklady na jednotlivé jedlá vynásobené počtom hostí sa sčítajú do celkovej ceny šeku (obrázok 5.1-9). A rovnako ako v stavebnictve sa recepty na varenie v reštaurácii nemusia meniť celé desaťročia. Na rozdiel od cien, kde sa náklady na ingrediencie môžu meniť každú hodinu.

Podobne ako majiteľ reštaurácie vynásobí náklady na každé jedlo počtom porcií a ľudí, aby určil celkové náklady na podujatie, manažér odhadovania nákladov sčíta náklady na všetky zložky projektu, aby získal kompletný odhad stavby.

Pre každú aktivitu v projekte sa teda určia jej konečné náklady (obr. 5.1-9), ktoré sa vynásobia objemom atribútu subjektu zodpovedajúceho tejto aktivite - získajú sa náklady skupín aktivít, z ktorých sa získajú konečné náklady celého projektu.



Obrázok 5.1-9 Konečný odhad sa vypočíta súčtom atribútu nákladov na prácu každého prvku podľa jeho atribútu rozsahu.

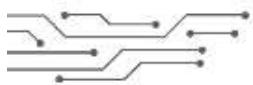
Celkové náklady na projekt (obrázok 5.1-8) poskytujú finančný obraz projektu, ktorý umožňuje zákazníkom, investorom alebo financujúcim organizáciám pochopiť celkový rozpočet a finančné zdroje potrebné na realizáciu projektu v daný deň, pričom sa zohľadňujú aktuálne ceny.

A ak sú procesy zostavovania zdrojových podkladov, kalkulácií a odhadov (procesné recepty) už vypracované, poloautomatizované a zdokonaľované desaťtisíce rokov a evidované na štátnej úrovni, potom automatické získavanie kvalitatívnych informácií o objeme a množstve prvkov pre poslednú fázu konečného odhadu - dnes zostáva úzkym miestom v procesoch všetkých kalkulácií nákladových a časových atribútov projektu a vôbec celkového rozpočtu projektu.

Tradičnou metódou výpočtu objemov boli po tisíce rokov manuálne metódy merania objemu a množstva pomocou plochých výkresov. S príchodom digitálneho veku spoločnosti zistili, že informácie o objeme a množstve sa teraz dajú automaticky extrahovať z geometrických údajov obsiahnutých v modeloch CAD, čím sa prevratne zmenili tisícročia staré spôsoby získavania kvantitatívnych údajov.

Moderné prístupy k odhadovaniu a oceňovaniu procesov zahŕňajú automatickú extrakciu objemových a kvantitatívnych atribútov z databáz CAD, ktoré možno nahrať a pripojiť k procesu kalkulácie nákladov, aby sa získali aktuálne objemy projektových skupín v ktorejkoľvek fáze návrhu

až po prevádzku.



## KAPITOLA 5.2.

# VÝPOČET MNOŽSTVA A AUTOMATICKÉ VYTVÁRANIE ODHADOV A HARMONOGRAMOV

### Prechod z 3D na 4D a 5D: využitie objemových a kvantitatívnych parametrov

Ked' máte v ruke kalkulačné tabuľky s opisanými procesmi prostredníctvom zdrojov (obr. 5.1-8), ďalším krokom je automatické získanie objemových alebo množstevných parametrov pre skupinu prvkov, ktoré sú potrebné pre výpočty a pre konečný odhad.

Objemové charakteristiky prvkov projektu - napr. stien alebo dosiek - možno automaticky získať z databáz CAD. Parametrické objekty vytvorené v programoch CAD sa pomocou geometrického jadra konvertujú na číselné hodnoty dĺžky, šírky, plochy, objemu a ďalších parametrov. Procesom získavania objemov na základe 3D geometrie sa budeme podrobnejšie zaoberať v ďalšej, šiestej časti (obr. 6.3-3), venovanej práci s CAD (BIM). Okrem objemov možno z databázy CAD-modelu získať aj počet podobných prvkov filtrovaním a zoskupovaním objektov podľa kategórií a vlastností. Tieto parametre, ktoré umožňujú zoskupovanie, sa stávajú základom na prepojenie prvkov projektu prostredníctvom výpočtov zdrojov s kalkuláciami, konečnými odhadmi a rozpočtom celého projektu.

Dátový model získaný z 3D (CAD) modelu je teda rozšírený o nové vrstvy parametrov, označované ako 4D a 5D. V nových vrstvách atribútov entity, 4D (čas) a 5D (náklady), sa ako zdroj hodnôt atribútov objemu entity používajú 3D geometrické údaje.

- **4D** je informačná vrstva parametrov, ktorá k 3D parametrom prvkov pridáva informácie o trvaní stavebných operácií. Tieto údaje sú nevyhnutné na plánovanie harmonogramov a riadenie časového harmonogramu projektu
- **5D** je ďalšou úrovňou rozšírenia dátového modelu, v ktorom sú prvky doplnené o nákladové charakteristiky. Tým sa ku geometrickým informáciám pridáva finančný aspekt: náklady na materiál, práce a zariadenia, čo umožňuje výpočet rozpočtu, analýzy ziskovosti a riadenie nákladov počas procesu výstavby.

Náklady a atribútové údaje 3D, 4D a 5D skupín projektových subjektov sú popísané podobným spôsobom ako výpočty v modulárnych ERP, PIMS -systémoch (alebo nástrojoch podobných Excelu) a používajú sa na automatické kalkulovanie nákladov a plánovanie rozpočtu jednotlivých skupín aj celého rozpočtu projektu.

### 5D atribúty a získavanie objemov atribútov z CAD

Pri príprave konečného odhadu stavebného projektu, ktorého príprave sme sa venovali v predchádzajúcich kapitolách (obrázok 5.1-8), sa objemové atribúty pre jednotlivé kategórie prvkov projektu zbierajú buď ručne, alebo sa získavajú zo špecifikácií objemových atribútov, ktoré poskytuje softvér CAD.

Tradičná manuálna metóda výpočtu množstiev zahŕňa analýzu výkresov, ktoré sa už tisíce rokov prezentujú ako čiary na papieri a posledných 30 rokov v digitálnych formátoch, ako je PDF (PLT) alebo DWG. Na základe odborných skúseností merajú množstvá potrebných prác a materiálov, často pomocou pravítka a uhlomeru. Táto metóda si vyžaduje značné úsilie a čas, ako aj osobitnú pozornosť venovanú detailom.

Určenie atribútov rozsahu týmto spôsobom môže trvať od niekoľkých dní do niekoľkých mesiacov v závislosti od rozsahu projektu. Okrem toho, keďže všetky merania a výpočty sa vykonávajú ručne, existuje riziko ľudskej chyby, ktorá môže viesť k nepresným údajom, čo následne ovplyvňuje chyby v odhade času a nákladov na projekt, za ktoré bude niest zodpovednosť celá spoločnosť.

Moderné metódy založené na využívaní databáz CAD výrazne zjednodušujú výpočet objemov. V modeloch CAD už geometria prvkov obsahuje atribúty objemu, ktoré možno automaticky vypočítať (prostredníctvom jadra geometrie (obr. 6.3-3)) a prezentovať alebo exportovať v tabuľkovej forme.

V takomto prípade oddelenie odhadu požiada projektanta CAD o údaje o množstevných a objemových charakteristikách prvkov projektu. Tieto údaje sa exportujú vo forme tabuľiek alebo sa priamo integrujú do databáz kalkulácie nákladov - či už ide o Excel, ERP alebo PMIS - systémy. Tento proces sa často nezačína formálou žiadosťou, ale krátkym dialógom medzi klientom (iniciátorom) a architektom odhadcom zo stavebnej alebo projektovej spoločnosti. Nižšie je uvedený zjednodušený príklad, ktorý ukazuje, ako sa z každodennej komunikácie vytvára štruktúrovaná tabuľka pre automatické výpočty (QTO):

- ⌚ Zákazník - "Chcem k budove pridať ďalšie poschodie v rovnakej konfigurácii ako prvé poschodie"
- ⌚ Architekt (CAD) - "Pridanie tretieho poschodia, konfigurácia je rovnaká ako na druhom poschodi".  
A po tejto správe odošle novú verziu projektu CAD odhadcovi.
- ⌚ Odhadca automaticky vykoná zoskupenie a výpočet (ERP, PMIS, Excel) - "spustím projekt cez tabuľku Excelu s pravidlami QTO (ERP, PMIS), získam objemy podľa kategórií pre novú podlahu a vygenerujem odhad"

Výsledkom je transformácia textového dialógu na tabuľkovú štruktúru s pravidlami zoskupovania:

Prvok	Kategória	Podlaha
Prekrytie	OST_Floors	3
Stípec	OST_StructuralColumns	3
schodisko	OST_Stairs	3

Po procese automatického zoskupenia CAD modelu od projektanta podľa pravidiel QTO odhadcu a automatického vynásobenia objemov výpočtom zdrojov (obr. 5.1-8) dostaneme nasledujúce výsledky, ktoré sa odošlú zákazníkovi:

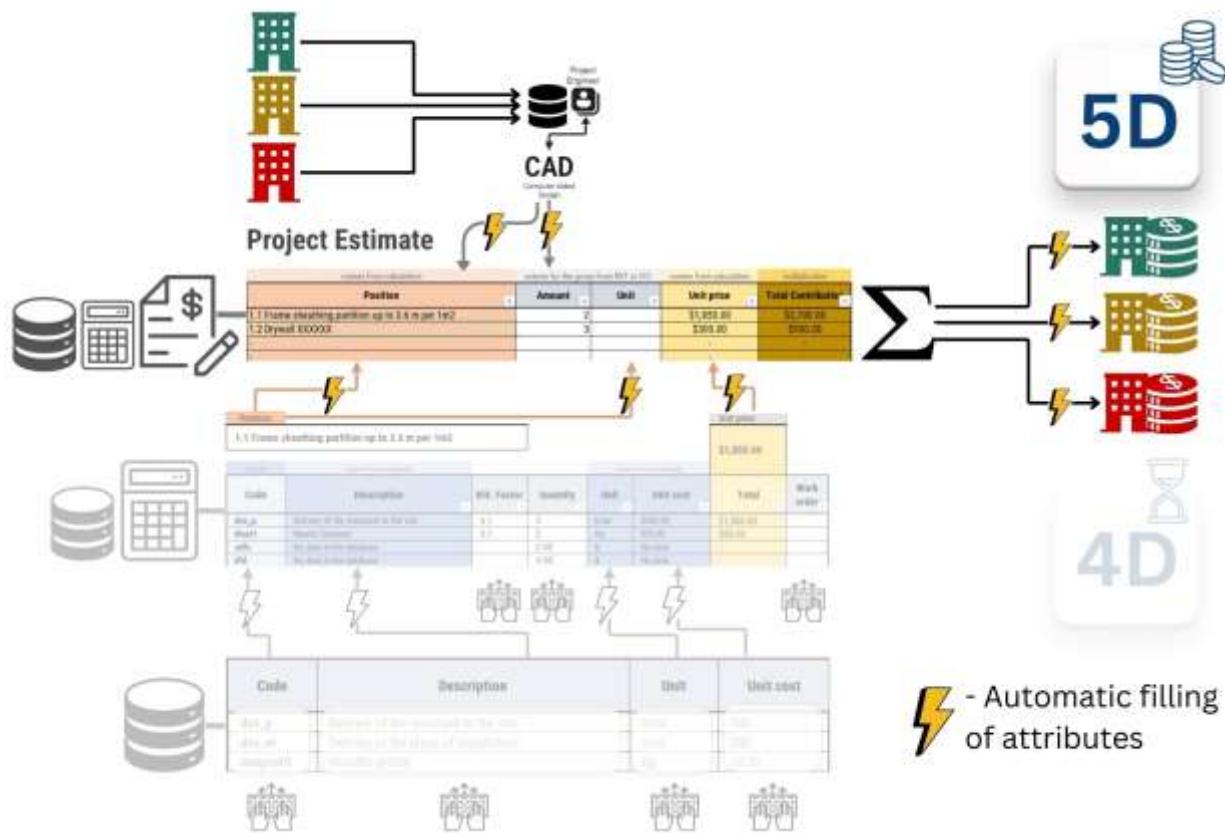
Prvok	Zväzok	Podlaha	Cena za jednotku.	Celkové náklady
Prekrytie	420 m <sup>2</sup>	3	150 €/m <sup>2</sup>	<b>63 000 €</b>
Stípec	4 ks.	3	2450 €/kus.	<b>9 800 €</b>
schodisko	2 ks.	3	4 300 €/kus.	<b>8 600 €</b>
<b>SPOLU:</b>	-	-	-	<b>81 400 €</b>

☞ Zákazník - "Ďakujem, je to dosť veľa, potrebujeme skrátiť niekoľko izieb". A cyklus sa opakuje.

Tento scenár sa môže opakovať mnohokrát, najmä vo fáze schvaľovania, keď zákazník očakáva okamžitú spätnú väzbu. V praxi sa však takéto procesy môžu pretiahnuť na niekoľko dní alebo dokonca týždňov. Dnes by vďaka zavedeniu automatického zoskupovania a pravidiel výpočtu mali byť činnosti, ktoré predtým zaberali značný čas, hotové v priebehu niekoľkých minút. Automatizované získavanie množstiev prostredníctvom pravidiel zoskupovania nielenže urýchľuje výpočty a odhady, ale minimalizáciou ľudského faktora znížuje pravdepodobnosť chýb a poskytuje transparentné a presné posúdenie nákladov na projekt.

Ak sa pri vytváraní 3D modelu v systéme CAD pôvodne zohľadnili požiadavky oddelenia odhadu (čo je v praxi zatial' zriedkavé) a názvy, identifikátory skupín prvkov a ich klasifikačné atribúty sú nastavené vo forme parametrov, ktoré sa zhodujú so štruktúrami skupín a tried odhadu, potom sa objemové atribúty môžu automaticky preniesť do systémov odhadu bez dodatočných transformácií.

Automatická extrakcia objemových atribútov z CAD vo forme tabuľiek špecifikácií umožňuje rýchlo získať aktuálne údaje o nákladoch na jednotlivé práce a projekt ako celok (obr. 5.2-1). Aktualizáciou iba súboru CAD s objemami projektu v procese výpočtu alebo v kalkulačnom systéme môže spoločnosť rýchlo prepočítať odhad so zohľadením najnovších zmien, čím sa zabezpečí vysoká presnosť a konzistentnosť všetkých následných výpočtov.



Obr. 5.2-1 Objemové atribúty z tabuľiek CAD alebo databáz sú automaticky vložené do odhadu, čo umožňuje okamžitý výpočet celkových nákladov na projekt.

S narastajúcou zložitosťou kapitálových projektov sa výpočet celého rozpočtu a analýza celkových nákladov projektov podľa takéhoto scenára (obrázok 5.2-1) - stáva klúčovým nástrojom pre informované rozhodovanie.

Podľa štúdie spoločnosti Accenture Creating More Value through Capital Projects (2024) [20] popredné spoločnosti aktívne integrujú dátovú analytiku do digitálnych iniciatív a využívajú historické informácie na predpovedanie a optimalizáciu výsledkov. Z výskumu vyplýva, že čoraz viac vlastníkov a prevádzkovateľov využíva analýzu veľkých dát na predpovedanie trhových trendov a posúdenie komerčnej životoschopnosti ešte pred začiatím projektovania. Dosahuje sa to analýzou dátových skladov z existujúceho portfólia projektov. Okrem toho 79% vlastníkov - prevádzkovateľov zavádza "robustnú" prediktívnu analýzu na posúdenie výkonnosti projektu a podporu prevádzkového rozhodovania v reálnom čase.

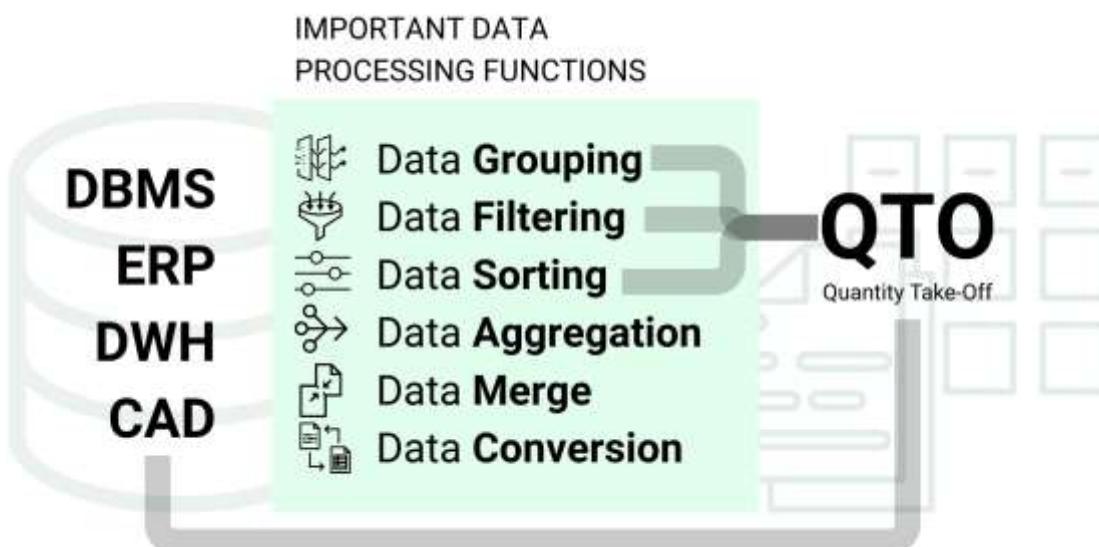
Moderné efektívne riadenie stavebných projektov je neoddeliteľne spojené so spracovaním a analýzou veľkého množstva informácií vo všetkých fázach projektovania a v procesoch, ktoré projektovaniu predchádzajú. Využívanie dátových skladov, výpočtov zdrojov, prediktívnych modelov a strojového učenia umožňuje nielen minimalizovať riziká pri výpočtoch, ale aj prijímať strategické rozhodnutia o financovaní projektu v počiatočných fázach projektovania. O dátových skladoch a prediktívnych modeloch, ktoré dopĺnia výpočty, budeme viac hovoriť v deviatej časti knihy.

Automatické získavanie objemových parametrov prvkov z projektov CAD, ktoré sú potrebné na prípravu odhadov, sa vykonáva pomocou nástrojov na zoskupovanie QTO (Quantity Take-Off). Nástroje QTO pracujú tak, že zoskupujú všetky objekty projektu podľa špeciálnych identifikátorov prvkov alebo parametrov atribútov prvkov, pričom využívajú špecifikácie a tabuľky vytvorené v databáze CAD.

## **QTO Quantity Take-Off: zoskupovanie údajov o projekte podľa atribútov**

QTO (Quantity Take-Off) v stavebníctve je proces získavania množstva prvkov potrebných na realizáciu projektu. V praxi je QTO často čiastočne manuálny proces, ktorý zahŕňa zber údajov z rôznych zdrojov: dokumentov PDF, výkresov DWG a digitálnych modelov CAD.

Pri práci s údajmi získanými z databáz CAD sa proces QTO realizuje ako postupnosť operácií filtrovania, triedenia, zoskupovania a agregácie. Prvky modelu sa vyberajú podľa parametrov tried, kategórií a typov, potom sa ich kvantitatívne atribúty - ako napríklad objem, plocha, dĺžka alebo množstvo - sumarizujú podľa logiky výpočtu (obr. 5.2-2).



Obr. 5.2-2 Zoskupovanie a filtrovanie údajov sú najobľúbenejšie funkcie, ktoré sa uplatňujú v databázach a dátových skladoch.

Proces QTO (filtrovanie a zoskupovanie) umožňuje systematizovať údaje, vytvoriť špecifikácie a pripraviť vstupné informácie na výpočet odhadov, nákupov a harmonogramov prác. Základom QTO je klasifikácia prvkov podľa typu meraných atribútov. Pre každý prvok alebo skupinu prvkov sa vyberie zodpovedajúci kvantitatívny parameter merania. Napríklad:

- **Atribút dĺžky** (obrubník - v metroch)
- **Atribút plochy** (sadrokartónové práce - v metroch štvorcových)
- **Atribút objemu** (betónové práce - v metroch kubických)
- **Atribút množstva** (okná - na kus)

Okrem objemových charakteristik vytvorených matematicky na základe geometrie sa vo výpočtoch po zoskupení QTO často používajú faktory prekročenia (obr. 5.2-12, napr. 1,1 na zohľadnenie 10% na logistiku a inštaláciu) - korekčné hodnoty, ktoré zohľadňujú straty, inštaláciu, skladovanie alebo prepravu. To umožňuje presnejšie predvídať skutočnú spotrebu materiálov a vyhnúť sa nedostatku aj nadmerným zásobám na stavbe.

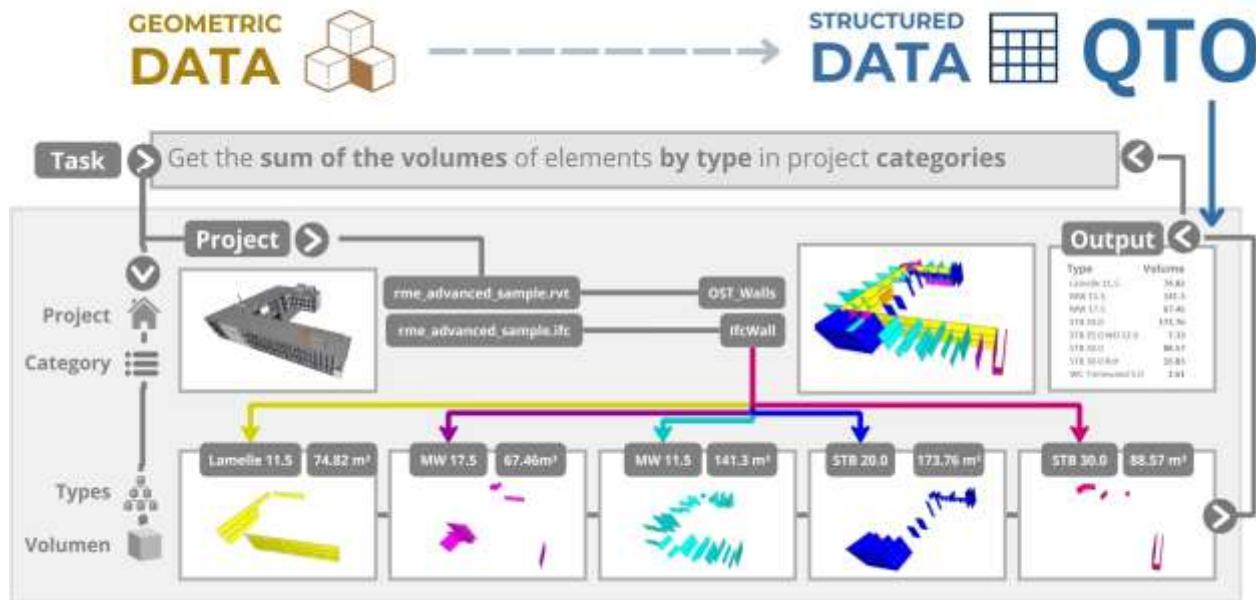
Automatizovaný proces odberu množstva (QTO) je nevyhnutný na vypracovanie presných výpočtov a odhadov, na zníženie ľudskej chyby v procesoch špecifikácie objemu a na zabranenie nadmernému alebo nedostatočnému objednávaniu materiálov.

Ako príklad procesu QTO si uveďme bežný prípad, keď je potrebné z databázy CAD zobraziť tabuľku - špecifikáciu objemov podľa typov prvkov pre určitú kategóriu, triedy prvkov. Zoskupme všetky prvky projektu podľa typu z kategórie steny projektu CAD a zosumarizujme atribúty objemov pre každý typ, aby sme výsledok prezentovali ako QTO tabuľku objemov (obr. 5.2-3).

V príklade typického projektu CAD (obr. 5.2-3) sú všetky prvky kategórie stien v databáze CAD zoskupené podľa typu steny, napríklad "Lamelle 11.5", "MW 11.5" a "STB 20.0", a majú presne definované atribúty objemu reprezentované v metrických kockách.

Cieľom manažéra, ktorý je na rozhraní medzi projektantmi a špecialistami na výpočty, je získať automatizovanú tabuľku objemov podľa typu prvku vo vybranej kategórii. Nielen pre konkrétny projekt, ale aj v univerzálnnej podobe použiteľnej pre iné projekty s podobnou štruktúrou modelu. Vďaka tomu je tento prístup škálovateľný a umožňuje opakované použitie údajov bez duplicitného úsilia.

Preč sú časy, keď sa skúsení projektanti a odhadcovia vyzbrojili pravítkom a starostlivo merali každú čiaru na papieri alebo v PDF -plánoch - tradícia, ktorá sa za posledné tisíce ročník nezmenila. S rozvojom 3D -modelovania, kde je teraz geometria každého prvku priamo prepojená s automaticky vypočítanými objemovými atribútmi, sa proces určovania objemov a QTO veličín zautomatizoval.



Obrázok 5.2-3 Získanie atribútov objemu a množstva QTO z projektu zahŕňa zoskupovanie a filtrovanie prvkov projektu.

V našom príklade je úlohou "vybrať kategóriu stien v projekte, zoskupiť všetky prvky podľa typu a prezentovať informácie o atribútoch objemu v štruktúrovanej tabuľkovej forme" tak, aby túto tabuľku mohli využívať desiatky ďalších špecialistov na výpočty kalkulácií, logistiky, harmonogramov prác a iných obchodných prípadov (obr. 6.1-3).

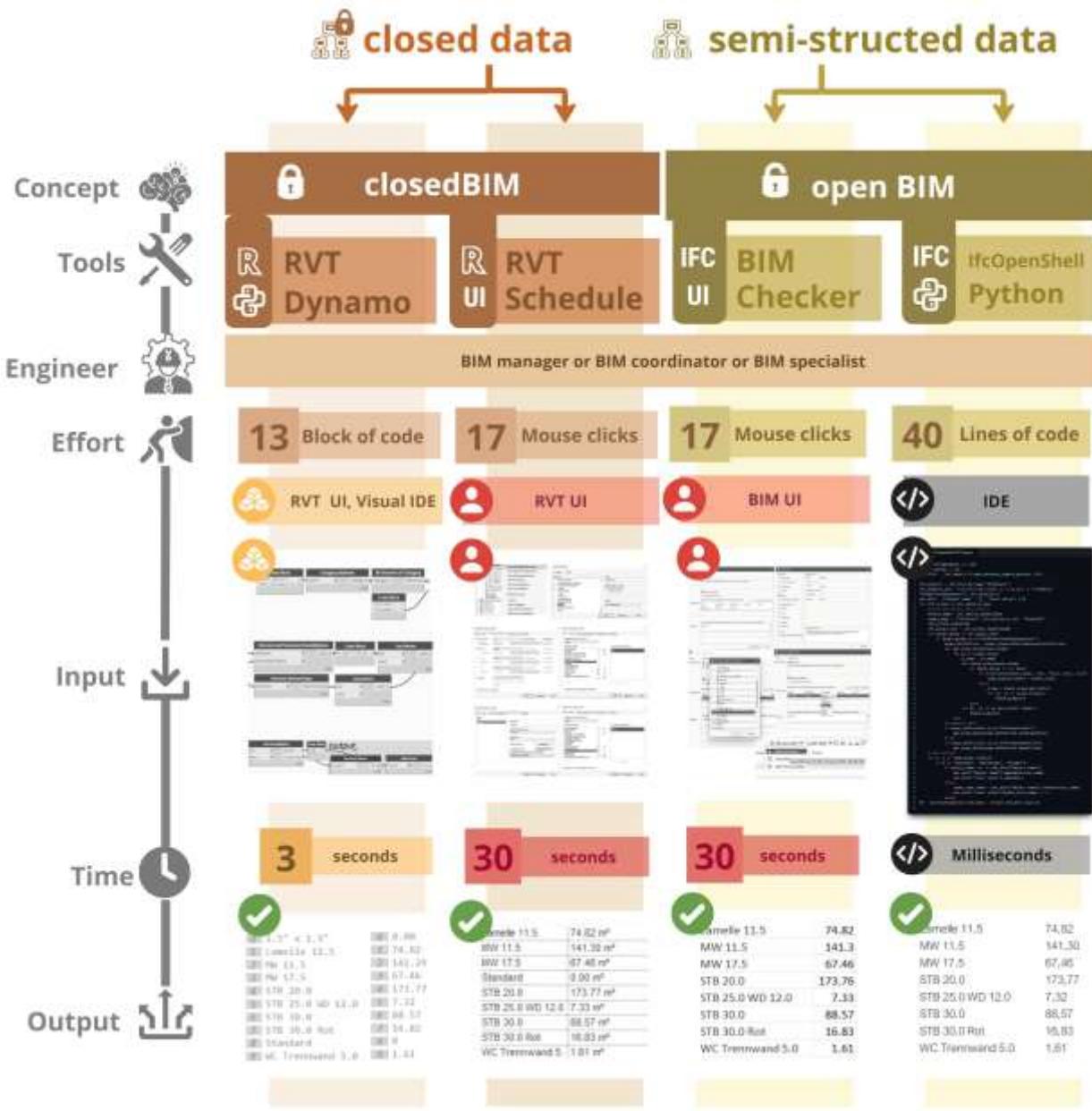
Vzhľadom na uzavretosť údajov CAD nie každý odborník dnes môže využívať priamy prístup do databázy CAD (dôvody a riešenia problému prístupu sú podrobne uvedené v šiestej časti knihy). Preto sa mnohí musia obrátiť na špecializované nástroje BIM založené na konceptoch otvoreného BIM a uzavretého BIM [63]. Pri práci so špecializovanými nástrojmi BIM alebo priamo v prostredí softvéru CAD sa tabuľka s výsledkami QTO (Quantity Take-Off) môže vytvárať rôznymi spôsobmi - podľa toho, či sa používa manuálne rozhranie alebo softvérová automatizácia.

Napríklad pomocou používateľského rozhrania softvéru CAD (BIM) stačí vykonať približne 17 akcií (kliknutí na tlačidlo), aby sa získala hotová tabuľka objemov (obr. 5.2-4). Používateľ však musí dobre poznať štruktúru modelu a funkcie softvéru CAD (BIM).

Ak sa automatizácia použije prostredníctvom programového kódu alebo prostredníctvom zásuvných modulov a nástrojov API v rámci programov CAD, zníži sa počet manuálnych krokov na získanie objemových tabuľiek, ale v závislosti od použitej knižnice alebo nástroja bude potrebné napísať 40 až 150 riadkov kódu:

- **IfcOpSh (otvorený BIM)** alebo **Dynamo IronPython (uzavretý BIM)** - umožňujú získať tabuľku QTO z formátu CAD alebo programu CAD len v ~40 riadkoch kódu.
- **IFC\_js (otvorený BIM)** - vyžaduje približne 150 riadkov kódu na extrakciu objemných atribútov z modelu IFC.

- **Rozhranie nástrojov CAD (BIM)** - umožňuje získať rovnaký výsledok ručne, na 17 kliknutí myšou.

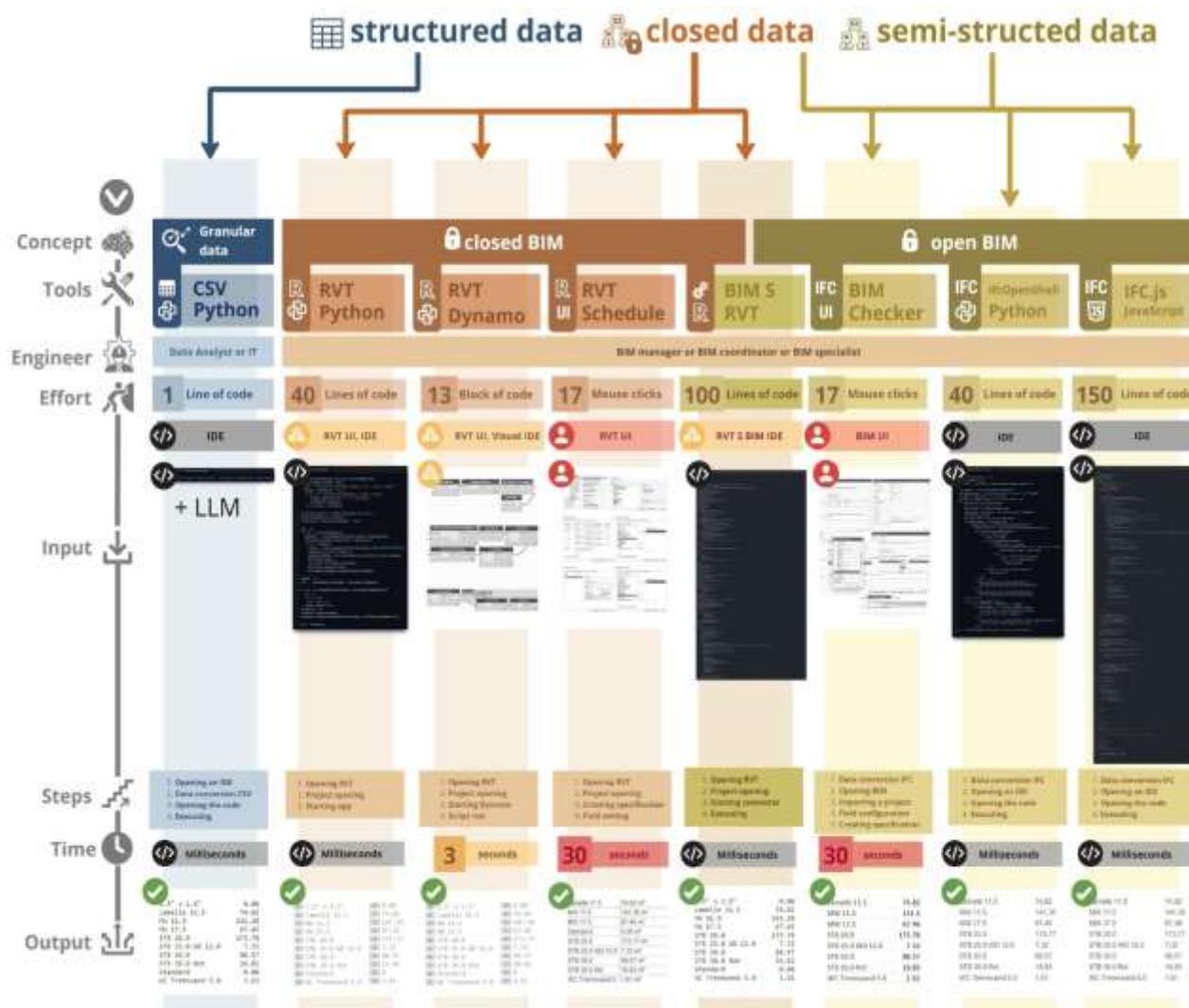


Obrázok 5.2-4 CAD (BIM) projektanti a manažéri, používajú 40 až 150 riadkov kódu alebo desiatky stlačení kláves na vytvorenie QTO tabuliek.

Výsledok je rovnaký - štruktúrovaná tabuľka s atribútmi objemu pre skupinu prvkov. Jediným rozdielom sú náklady na prácu a úroveň technických znalostí potrebných pre používateľa (obr. 5.2-4). Moderné nástroje v porovnaní s manuálnym zberom objemov výrazne urýchľujú proces QTO a znižujú pravdepodobnosť chýb. Umožňujú získavať údaje priamo z modelu projektu, čím sa eliminuje potreba ručného prepočítavania objemov z výkresov, ako sa to robilo v minulosti.

Bez ohľadu na použitú metódu - či už ide o otvorený BIM alebo uzavretý BIM - je možné získať identickú QTO - tabuľku s objemami prvkov projektu (obr. 5.2-4). Pri práci s údajmi projektu v CAD - (BIM-) koncepcii sú však používateľia závislí od špecializovaných nástrojov a API poskytovaných dodávateľmi (obr. 3.2-13). To vytvára ďalšie vrstvy závislosti a vyžaduje si osvojenie jedinečných dátových schém, pričom obmedzuje priamy prístup k údajom.

Vzhľadom na uzavretý charakter údajov CAD komplikuje získavanie QTO tabuľiek a ďalších parametrov automatizáciu výpočtov a integráciu s externými systémami. Použitím nástrojov na priamy prístup k databázam a prenesením CAD -údajov projektu pomocou nástrojov reverzného inžinierstva do otvoreného formátu štruktúrovaného dátového rámca (obr. 4.1-13) možno získať identickú QTO tabuľku len jedným riadkom kódu (obr. 5.2-5 - variant s granulovanými údajmi).



Obrázok 5.2-5 Rôzne nástroje poskytujú rovnaké výsledky v podobe atribútových tabuľiek subjektov projektu, ale s rôznymi nákladmi na prácu.

Pri použití otvorených štruktúrovaných údajov z projektov CAD, ako je uvedené v kapitole "Konverzia údajov CAD (BIM) do štruktúrovanej podoby", sa výrazne zjednodušuje proces zoskupovania, QTO.

Prístupy založené na využívaní otvorených štruktúrovaných údajov alebo priameho prístupu k databázam modelov CAD sú osloboodené od marketingových obmedzení spojených so skratkou BIM. Spoliehajú sa na osvedčené nástroje dlhodobo používané v iných odvetviach (obr. 7.3-10. Proces ETL).

Podľa štúdie spoločnosti McKinsey "Open Data: Uvoľnite inovácie a produktivitu pomocou prúdenia informácií" [102] uskutočnej v roku 2013 by využívanie otvorených údajov mohlo vytvoriť príležitosť na úspory vo výške 30 až 50 miliárd USD ročne v oblasti projektovania, inžinierstva, obstarávania a výstavby zariadení na výrobu energie. To znamená 15-percentnú úsporu investičných nákladov na výstavbu.

Práca s otvorenými štruktúrovanými (granulovanými) údajmi zjednodušuje vyhľadávanie a spracovanie informácií, znížuje závislosť od špecializovaných platform BIM a otvára cestu k automatizácii bez potreby používať proprietárne systémy alebo parametrické a zložité dátové modely z formátov CAD.

## Automatizácia QTO pomocou LLM a štruktúrovaných údajov

Prevod neštruktúrovaných údajov do štruktúrovanej podoby výrazne zvyšuje efektívnosť rôznych procesov: zjednodušuje spracovanie údajov (obr. 4.1-1, obr. 4.1-2) a urýchľuje proces validácie tým, že požiadavky sú jasné a prehľadné, ako sme už uviedli v predchádzajúcich kapitolách. Podobne prevod údajov CAD (BIM) do štruktúrovanej otvorenej formy (obr. 4.1-12, obr. 4.1-13) uľahčuje proces zoskupovania atribútov a proces QTO.

Tabuľka atribútov QTO má štruktúrovanú podobu, takže pri použití štruktúrovaných údajov CAD pracujeme s jedným dátovým modelom (obrázok 5.2-5), čím odpadá potreba konverzie a prevodu dátových modelov projektu a pravidiel zoskupovania na spoločného menovateľa. To nám umožňuje zoskupovať údaje podľa jedného alebo viacerých atribútov len pomocou jedného riadku kódu. Naproti tomu v otvorenom BIM a uzavretom BIM, kde sú údaje uložené v pološtruktúrovaných, parametrických alebo uzavretých formátoch, si spracovanie vyžaduje desiatky alebo dokonca stovky riadkov kódu a použitie API na interakciu s geometriou a atribútmi informácií.

- ❷ Príklad zoskupenia štruktúrovaného projektu QTO podľa jedného atribútu. Textový dotaz v ľubovoľnej chatovej miestnosti LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akejkoľvek inej):

Mám CAD -projekt ako DataFrame - prosím, filtrovajte údaje o projekte, aby ste získali položky s parametrom "Type", ktorý obsahuje iba hodnotu "Type 1" ↴

- Odpoveď LLM bude s veľkou pravdepodobnosťou vo forme kódu Pythonu s použitím Pandas :



Obr. 5.2-6 Jeden riadok kódu napísaný pomocou LLM , umožňuje zoskupiť celý projekt CAD podľa atribútu "Type" a získať požadovanú skupinu prvkov.

Vďaka jednoduchej štruktúre 2D DataFrame nemusíme vysvetľovať schému a dátový model LLM, čo skracuje kroky výkladu a urýchľuje tvorbu finálnych riešení. Predtým si písanie aj jednoduchého kódu vyžadovalo učenie sa programovacích jazykov, ale teraz nám moderné jazykové modely (LLM) umožňujú automaticky preložiť logiku procesu do kódu pri práci so štruktúrovanými údajmi pomocou textových dotazov.

Automatizácia a jazykové modely LLM môžu úplne eliminovať potrebu odborníkov pracujúcich so zoskupovaním a spracovaním údajov CAD (BIM) učiť sa programovacie jazyky alebo nástroje BIM tým, že poskytujú možnosť riešiť problémy pomocou textových dotazov.

Ten istý dotaz - zoskupenie všetkých prvkov projektu z kategórie "steny" a výpočet objemov pre každý typ (obr. 5.2-5) - ktorý v prostredí CAD (BIM) vyžaduje 17 kliknutí v rozhraní alebo napísanie 40 riadkov kódu, v otvorených nástrojoch na spracovanie údajov (napr. SQL alebo Pandas) vyzerá ako jednoduchý a intuitívny dotaz:

- Pomocou jediného riadku v programe Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

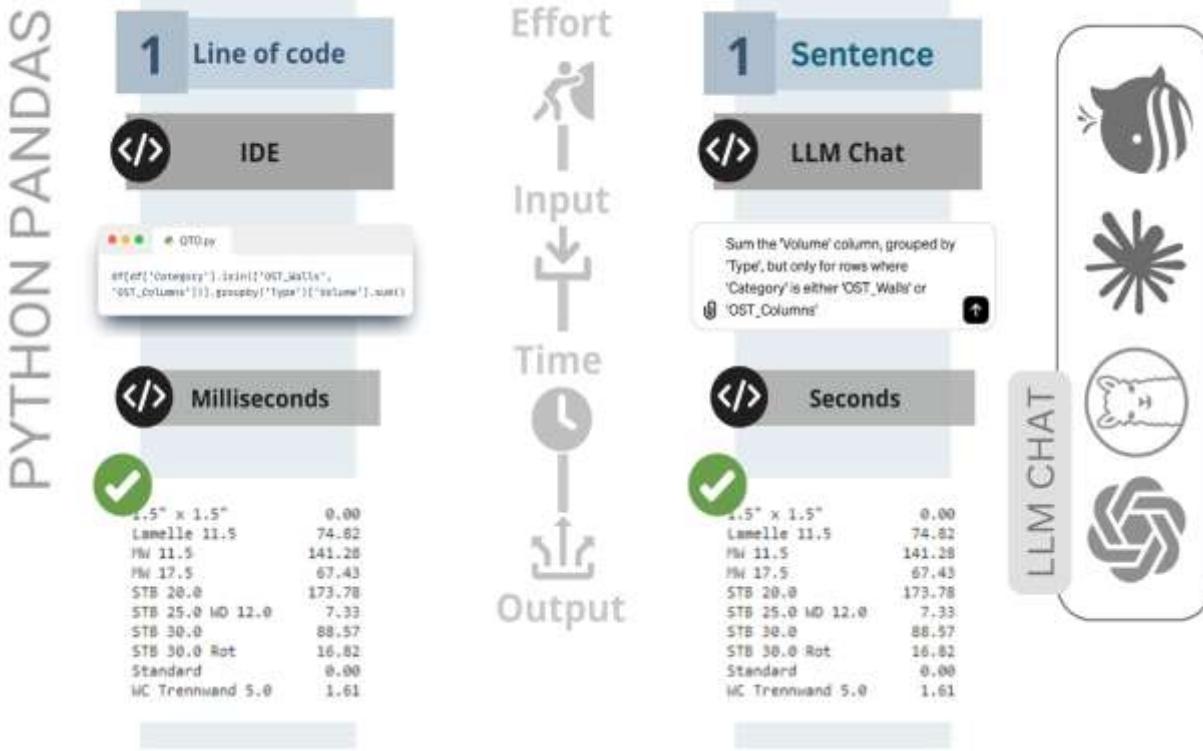
Dekódovanie kódu: z df (DataFrame) vyberte prvky, ktorých atribút-stĺpec "Category", má hodnoty "OST\_Walls", zoskupte všetky získané prvky podľa atribútu-stĺpca "Type" a pre získanú skupinu prvkov zráťajte atribút "Volume".

- Zoskupenie štruktúrovaného projektu získaného z CAD pomocou SQL:

```
SELECT Typ, SUM(Objem) AS Celkový objem
FROM prvky
WHERE Category = 'OST_Walls'
GROUP BY Type;
```

- Pomocou LLM môžeme zapísať požiadavku na zoskupenie do databázy projektu ako jednoduchý textový odkaz - výzvu (obr. 5.2-7):

Pre dátový rámec projektu zoskupte položky podľa parametra "Typ", ale len pre položky s parametrom "Kategória" rovnajúcim sa "OST\_Steny" alebo "OST\_Stĺpce" a pre výslednú zosumarizujte parameter stĺpca "Objem" ↴

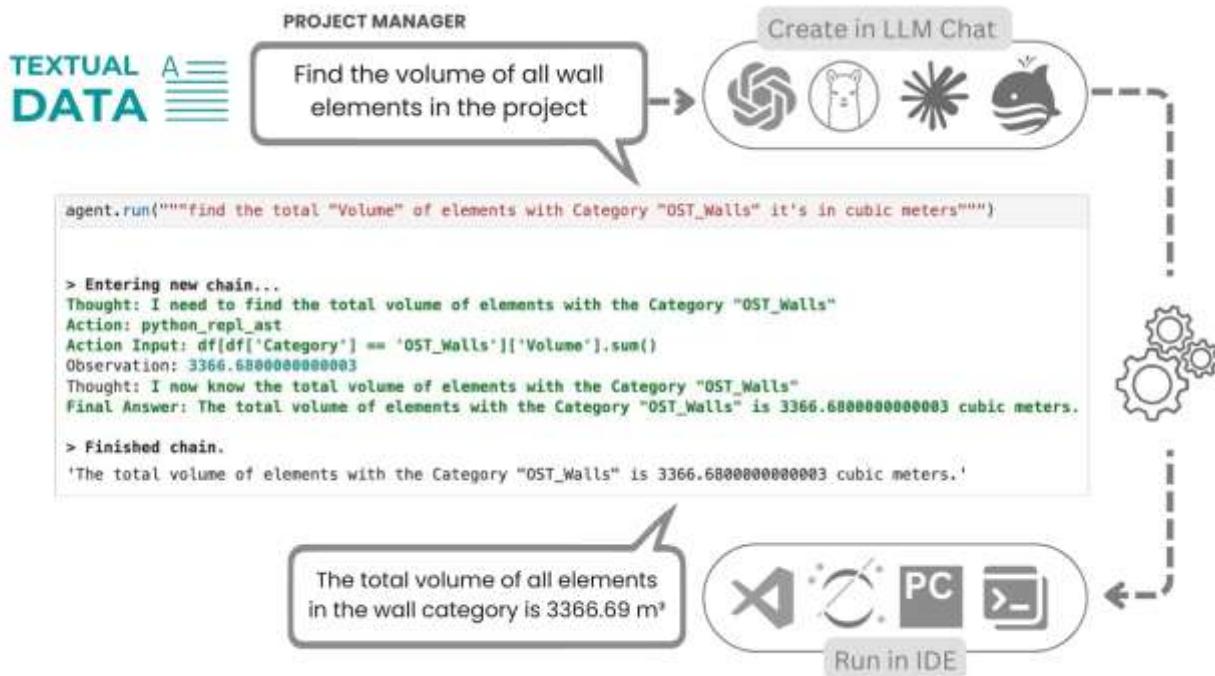


Obrázok 5.2-7 Pomocou SQL, Pandas a LLM je teraz možné automatizovať spracovanie údajov pomocou niekoľkých riadkov kódu a textových dotazov.

Získavanie QTO z údajov CAD pomocou nástrojov LLM (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), dramaticky mení tradičné metódy získavania atribútových informácií, kvantitatívnych a objemových údajov pre jednotlivé objekty a skupiny objektov.

Teraz aj projektoví manažéri, kalkulanti alebo špecialisti na logistiku, ktorí nemajú hlboké znalosti o projektovaní a nemajú špecializovaný softvér CAD - (BIM-) dodávateľov, ktorí majú prístup k databáze CAD, môžu získať celkový objem prvkov kategórie stien alebo iných objektov v priebehu niekoľkých sekúnd, a to jednoduchým napísaním alebo nadiktovaním dotazu.

V textových dotazoch (obr. 5.2-8) agent LLM modelu spracúva požiadavku používateľa na aplikáciu určitej funkcie na jeden alebo viac parametrov - stĺpcov tabuľky. Výsledkom je, že používateľ v komunikácii s LLM dostane buď nový stĺpec - parameter s novými hodnotami, alebo jednu konkrétnu hodnotu po zoskupení



Obr. 5.2-8 Model LLM, ktorý pracuje so štruktúrovanými údajmi, z kontextu textového dotazu pochopí, na aké zoskupenie a atribúty sa používateľ pýta.

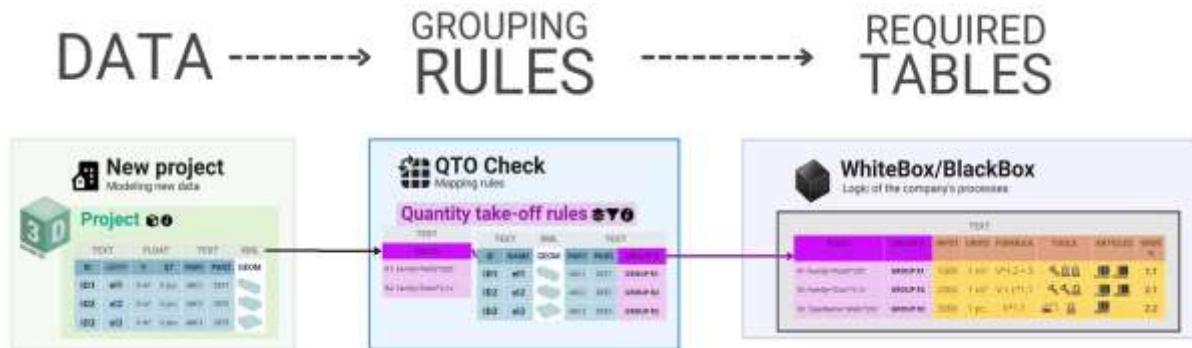
Ak je potrebné získať množstvá len pre jednu skupinu prvkov, stačí vykonať jednoduchý dotaz QTO (obrázok 5.2-7) na údaje modelu CAD. Pri výpočte rozpočtu alebo odhadu pre celý projekt pozostávajúci z mnohých skupín prvkov je však často potrebné získať množstvá pre všetky typy prvkov (tryed0, pričom každá kategória prvkov sa spracúva samostatne - so zoskupením podľa príslušných atribútov).

V praxi odhadcov a posudzovateľov sa používajú jednotlivé pravidlá zoskupovania a výpočtu pre rôzne typy objektov. Napríklad okná sa zvyčajne zoskupujú podľa podlaží alebo zón (parameter zoskupenia - atribút Úroveň, Miestnosti) a steny podľa materiálu alebo typu konštrukcie (parameter Materiál, Typ). Na automatizáciu procesu zoskupovania sú takéto pravidlá vopred popísané na stránke vo forme

tabuľiek pravidiel zoskupovania. Tieto tabuľky fungujú ako konfiguračné šablóny, ktoré definujú, ktoré atribúty sa majú použiť pri výpočtoch pre každú skupinu prvkov v projekte.

## QTO výpočet celého projektu pomocou skupinových pravidiel z tabuľky Excel

V reálnych stavebných projektoch je často potrebné vykonať agregáciu podľa viacerých atribútov súčasne v rámci jednej skupiny prvkov. Napríklad pri práci s kategóriou "Okná" (kde atribút Category obsahuje hodnoty ako OST\_Windows alebo IfcWindows) možno prvky zoskupovať nielen podľa typu - napríklad podľa hodnoty v poli Type Name alebo Type - ale aj podľa ďalších charakteristík, ako je napríklad úroveň tepelnej vodivosti uvedená v príslušnom atribúte. Toto viacozmerné zoskupovanie umožňuje presnejšie výsledky pre konkrétnu skupinu. Podobne pri výpočte kategórií stien alebo podlág možno ako filtre alebo kritériá zoskupovania použiť ľubovoľné kombinácie atribútov - napríklad materiál, úroveň, podlaha, požiarna odolnosť a ďalšie parametre (obrázok 5.2-9).



Obrázok 5.2-9 Pre každú skupinu alebo kategóriu subjektov v projekte existuje iný vzorec na zoskupenie pozostávajúci z jedného alebo viacerých kritérií.

Proces definovania takýchto pravidiel zoskupovania je podobný procesu tvorby požiadaviek na údaje opísanému v kapitole "Tvorba požiadaviek a kontrola kvality údajov" (obrázok 4.4-5), kde sme sa podrobne venovali práci s dátovými modelmi. Tieto pravidlá zoskupovania a výpočtu zabezpečujú presnosť a relevantnosť výsledkov automatického výpočtu celkových atribútov množstva alebo atribútov objemu kategórie entít, pričom sa zohľadňujú všetky potrebné podmienky, ktoré sa musia zohľadniť pri výpočtoch a kalkuláciach.

- ❷ Nasledujúca ukážka kódu filtrouje tabuľku projektov tak, aby výsledný súbor údajov obsahoval len entity, v ktorých stĺpec atribútov "Category" obsahuje hodnoty "OST\_Windows" alebo "IfcWindows" a zároveň stĺpec atribútov "Type" obsahuje hodnotu "Type 1":

Mám projekt DataFrame - filtrovanie údajov tak, aby v datasete zostali len položky, ktoré majú atribút "Category" obsahujúci hodnoty "OST\_Windows" alebo "IfcWindows" a zároveň atribút Type obsahuje hodnotu "Type 1" ↴

💡 Odpoveď LLM :



```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

Obrázok 5.2-10 Jediný riadok kódu, podobný vzorcu programu Excel , umožňuje zoskupiť všetky entity projektu podľa niekoľkých atribútov.

Výsledný kód (obr. 5.2-10) po preklade údajov CAD v štruktúrovaných otvorených formátoch (obr. 4.1-13) možno spustiť v jednom z populárnych IDE (integrovaných vývojových prostredí), ktoré sme spomenuli vyššie, v offline režime: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo populárne online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 Ak chceme získať entity projektu vo formulári QTO DataFrame v kategórii "Windows" len s konkrétnou hodnotou tepelnej vodivosti, môžeme použiť nasledujúci dotaz do LLM:

Mám projekt DataFrame - filtrovanie údajov tak, aby v súbore údajov zostali len záznamy s hodnotou "Category" obsahujúcou hodnoty "OST\_Windows" alebo "IfcWindows" a zároveň aby stípec ThermalConductivity mal hodnotu 0,✉

💡 Odpoveď LLM :



```
group.py
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

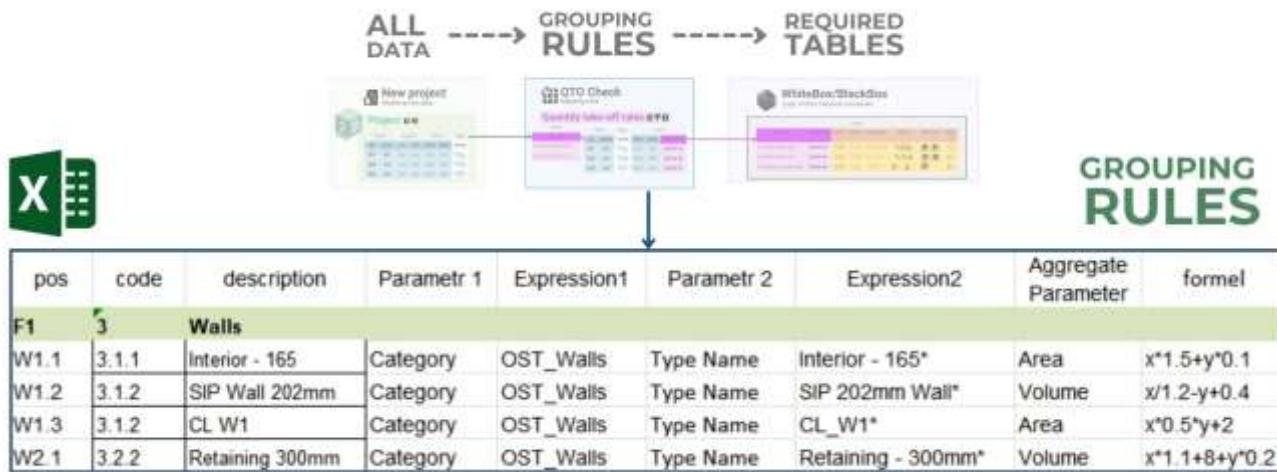
Obrázok 5.2-11 Mimoriadne jednoduchý dotazovací jazyk Pandas Python umožňuje spúštať QTO pre ľubovoľný počet projektov súčasne.

V odpovedi priatej od LLM (obr. 5.2-11) sa používa logická podmienka "&" na kombináciu dvoch kritérií: hodnota tepelnej vodivosti a príslušnosť k jednej z dvoch kategórií. Metóda "isin" kontroluje, či sa hodnota atribútu-stĺpca "Category" (kategória) nachádza v poskytnutom zozname.

V projektoch s veľkým počtom skupín prvkov s rôznou logikou zoskupovania - pre každú kategóriu subjektov projektu (napr.: okná, dvere, dosky) by sa mali stanoviť individuálne pravidlá zoskupovania, ktoré môžu zahŕňať dodatočné koeficienty alebo vzorce výpočtu celkových atribútov. Tieto vzorce (obrázok 5.2-12 atrubútový "formel", napr. x-hodnota množstva a y-objem skupiny) a koeficienty zohľadňujú jedinečné vlastnosti každej skupiny, napr:

- % prírastky k objemu materiálu na zohľadnenie prekročenia
- pevné dodatočné množstvo materiálu
- úpravy súvisiace s možnými rizikami a chybami vo výpočtoch vo forme vzorcov

Po sformulovaní pravidiel filtrovania a zoskupovania vo forme vzorcov parametrov pre každú kategóriu položiek ich možno uložiť ako tabuľku po riadkoch - napríklad vo formáte Excel (obrázok 5.2-12). Uložením týchto pravidiel v štruktúrovanej forme možno proces extrakcie, filtrovania a zoskupovania údajov o projekte plne automatizovať. Namesto ručného písania mnohých samostatných dotazov systém jednoducho prečíta tabuľku parametrov a aplikuje príslušné pravidlá na model (celkový rámec údajov o projekte (obr. 4.1-13)), čím vytvorí konečné tabuľky QTO pre každú kategóriu prvkov projektu.



Obrázok 5.2-12 Tabuľka zoskupenia atrubútov QTO stanovuje pravidlá pre zoskupenie prvkov projektu, pričom poskytuje presný celkový počet a rozsah pre každú kategóriu.

Zozbierané pravidlá vám umožnia zoskupiť celý projekt a vykonať všetky potrebné výpočty vrátane úprav atrubútov objemu. Výsledkom je, že objemy sa privedú do "skutočného objemu", ktorý sa používa na výpočty a kalkulácie, a nie tie, ktoré boli pôvodne vo fáze návrhu v modeli CAD.

V procese automatického vytvárania objemových tabuľiek QTO pre celý projekt musí aplikácia prejsť všetky kategórie tabuľky pravidiel zoskupovania, prevziať atribúty zoskupovania, zoskupiť podľa nich všetky prvky projektu a agregovať atribút objemu pre túto skupinu, pričom ho navyše vynásobí spresňujúcim faktorom alebo koeficientom.

Požiadajme LLM, aby nám napísal kód pre takéto riešenie, kde by kód musel načítať dve tabuľky - tabuľku pravidiel zoskupovania (obrázok 5.2-12) a tabuľku údajov zo samotného projektu (obrázok 4.1-13), a potom aplikovať pravidlá zoskupovania, zoskupiť položky podľa daných pravidiel, vypočítať agregované hodnoty a výsledky uložiť do nového súboru Excel.

✉ Pošlite textovú požiadavku na chat LLM:

Potrebujem kód na načítanie údajov o projekte zo súboru 'basic\_sample\_project.xlsx' a potom pravidlá zo súboru 'Grouping\_rules\_QTO.xlsx', pomocou ktorých sa zoskupia všetky údaje podľa 'Parameter 1' a 'Parameter 2', agregujú sa 'Aggregate Parameter', filtrovú sa podľa 'Expression2', vykonajú sa výpočty z 'Formel1' a uloží sa tabuľka QTO do súboru 'QTQ\_table2.xlsx' ↵

☒ Odpoveď LLM :

```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 df.apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d*)')|0], fillna(0)))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     group = group.agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formula"], {'x': group['sum'], 'y': group['count']})
21     return group.reset_index()
22
23 # Apply the function to each row in the grouping rules dataframe and concatenate the results
24 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
25
26 # Save the processed data to an Excel file using the OpenPyXL engine
27 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Obrázok 5.2-13 Kód jazyka Python Pandas spracováva údaje o návrhu z projektu CAD a zoskupuje ich podľa pravidiel definovaných v programe Excel .

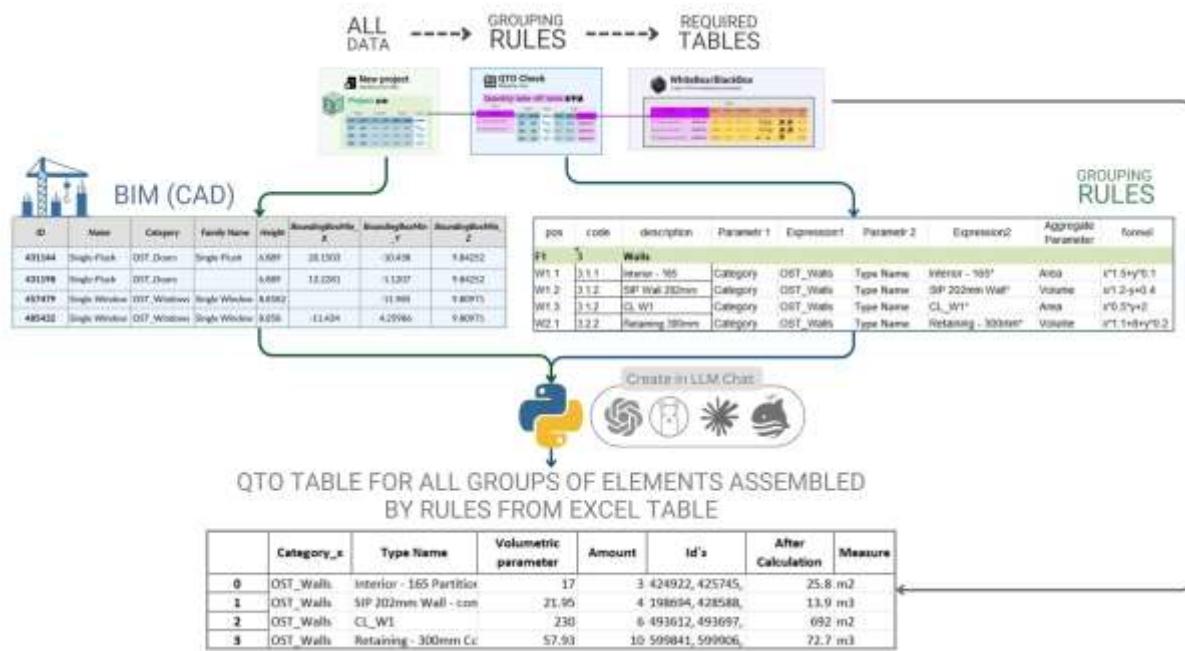
Konečným výsledkom vykonania kódu (obr. 5.2-13) bude tabuľka skupiny entít, ktorá obsahuje nielen zovšeobecnené atribúty objemu z pôvodného modelu CAD- (BIM-), ale aj nový atribút reálneho objemu, ktorý zohľadňuje všetky požiadavky na správne vytvorenie výpočtov a odhadov (príklad obr. 5.2-14).

	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,		25.8 m2
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,		13.9 m3
2	OST_Walls	CL_W1		230	6 493612, 493697,		692 m2
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,		72.7 m3

Obrázok 5.2-14 Atribút "Po výpočte" sa pridá do súhrnej tabuľky po vykonaní kódu, ktorý automaticky vypočíta skutočný objem.

Výsledný kód (obr. 5.2-13) možno spustiť v niektorom z populárnych IDE (ktoré sme spomenuli vyššie) a aplikovať ho na ľubovoľný počet existujúcich alebo nových prichádzajúcich projektov (RVT, IFC, DWG,

NWS, DGN atď.), či už ide o niekoľko projektov alebo možno stovky projektov v rôznych formátoch v štruktúrovanej podobe (obr. 5.2-15).



Obrázok 5.2-15 Proces automatického zoskupovania stavebných údajov spája údaje BIM (CAD) s tabuľkami QTO prostredníctvom pravidiel z tabuľky Excel.

Prispôsobený a parametrizovaný proces zberu objemových údajov (obrázok 5.2-15) umožňuje plne automatizovaný zber údajov o kvantitatívnych atribútoch a objemoch prvkov projektu pre ďalšiu prácu s nimi vrátane odhadu nákladov, logistiky, harmonogramov prác a výpočtu uhlíkovej stopy a ďalších analytických úloh.

Po osvojení si nástrojov, ktoré nám umožňujú jednoducho organizovať a zoskupovať skupiny prvkov projektu podľa určitých atribútov, sme teraz pripravení integrovať zoskupené a vyfiltrované projekty do rôznych firemných výpočtov a obchodných scenárov.



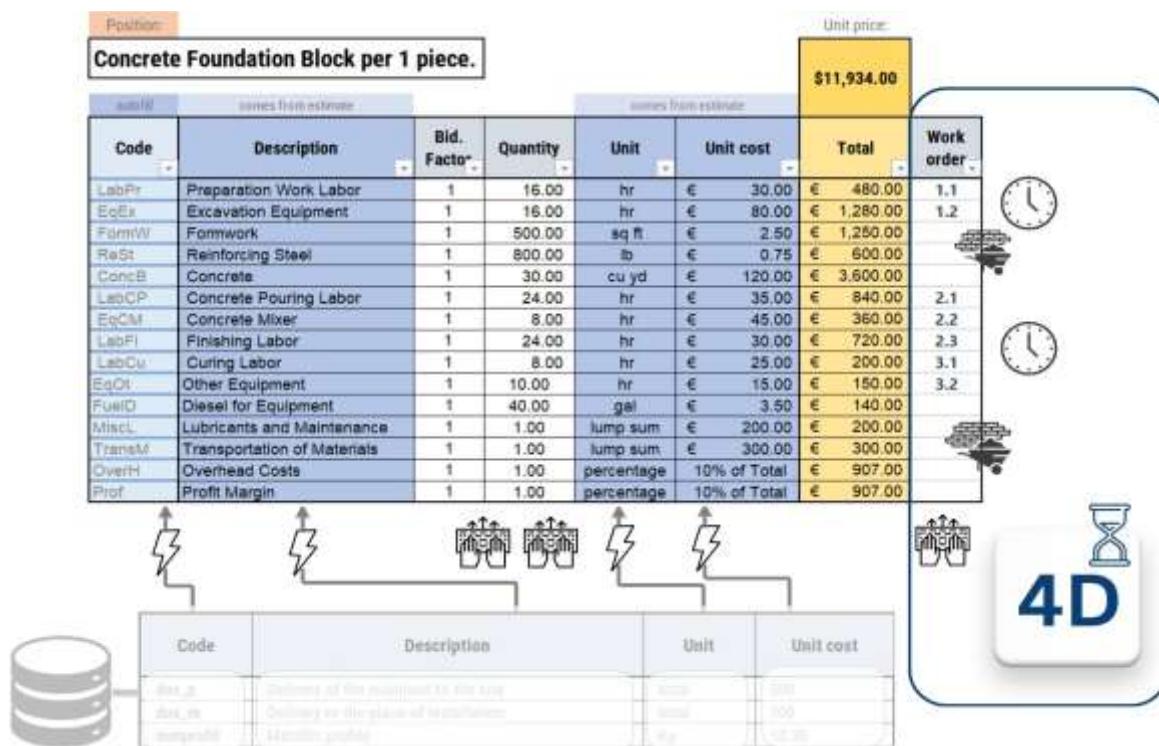
## KAPITOLA 5.3.

### 4D, 6D -8D A VÝPOČET EMISIÍ OXIDU UHLIČITÉHO CO<sub>2</sub>

#### 4D model: integrácia času do stavebných odhadov

Okrem kalkulácie nákladov je jednou z kľúčových aplikácií projektových údajov v stavebníctve určenie časových parametrov - pre jednotlivé stavebné operácie aj pre celý projekt. Metóda odhadu na základe zdrojov a súvisiaca kalkulačná databáza, ktorá bola podrobne rozobraná v predchádzajúcej kapitole "Výpočty a odhady pre stavebné projekty", sa často používa ako základ pre automatizovaný výpočet času a vytvorenie harmonogramu realizácie prác.

Prístup založený na zdrojoch zohľadňuje nielen materiálové náklady, ale aj časové zdroje. Pri kalkulácii nákladov možno každému procesu priradiť atribút pracovného príkazu (obrázok 5.3-1 - Parameter pracovného príkazu) a množstvo času a nákladov spojených s vykonaním tohto procesu. Tieto parametre sú dôležité najmä pri opise operácií, ktoré nemajú pevnú trhovú cenu a nedajú sa priamo nakúpiť - napríklad použitie stavebných zariadení, zamestnanie robotníkov alebo logistické procesy (ktoré sa zvyčajne bežne vyjadrujú v hodinách). V takýchto prípadoch náklady neurčuje oddelenie verejného obstarávania, ale priamo realizačná spoločnosť na základe interných noriem alebo výrobných sadzieb (obrázok 5.3-1).



Obr. 5.3-1 Výpočty práce pri metóde odhadu založenej na zdrojoch zahŕňajú časový rozvrh pracovných hodín.

Výpočty na úrovni nákladov teda zahŕňajú nielen náklady na palivo a materiál (nákupné náklady), ale aj

čas strávený vodičmi, technikmi a pomocnými pracovníkmi na mieste. V uvedenom príklade (obr. 5.3-1) je tabuľka nákladov kalkuláciou nákladov na inštaláciu základového bloku, ktorá zahrňa jednotlivé zložky práce, ako je príprava, inštalácia rámu a betonáž, ako aj potrebný materiál a prácu. Jednotlivé operácie, ako napríklad prípravné práce, však nemusia mať žiadne materiálové náklady, ale môžu obsahovať významné dočasné náklady na prácu vyjadrené v človekohodinách.

Na plánovanie postupnosti prác (pre harmonogram prác) na stavenisku sa do výpočtovej tabuľky ručne pridá atribút "Pracovná zákazka" (obr. 5.3-1). Tento atribút sa uvádza v dodatočnom stĺpci len pre položky, ktorých merná jednotka je vyjadrená v čase (hodina, deň). Tento atribút dopĺňa kód práce, popis, množstvo, mernú jednotku (parameter "Jednotka") a náklady. Číselné poradie (parameter "Poradie prác") činností umožňuje stanoviť poradie, v ktorom sa úlohy na stavbe vykonávajú, a použiť ho pri plánovaní.

## **Harmonogram výstavby a jeho automatizácia na základe údajov o nákladoch**

Harmonogram výstavby predstavuje vizuálne znázornenie plánu prác a procesov, ktoré sa majú vykonať v rámci realizácie projektu. Vytvára sa na základe podrobnych výpočtov zdrojov (obr. 5.3-1), kde sú okrem nákladov na zdroje naplánované aj jednotlivé úlohy - práce podľa času a poradia.

Na rozdiel od spriemerovaných prístupov, pri ktorých sú časové odhady založené na typických hodinách pre inštaláciu materiálov alebo zariadení, pri metóde založenej na zdrojoch sa plánovanie zakladá na skutočných údajoch v kalkulácii nákladov. Každá položka odhadu týkajúca sa mzdových nákladov vychádza z použitého kalendára, ktorý zohľadňuje skutočné podmienky využívania zdrojov počas pracovného obdobia. Úprava produktívnych hodín prostredníctvom koeficientov na úrovni kalkulácie nákladov (obr. 5.3-1 parameter "Bid. koeficient"), umožňuje zohľadniť rozdiely v produktivite a sezónne zvláštnosti ovplyvňujúce časový harmonogram práce.

Na určenie dátumov začiatku a konca procesu pre harmonogram výstavby v Ganttovom grafe vezmemme hodnoty atribútu množstva času pre každú položku z kalkulácie nákladov na základové bloky a vynásobíme ich počtom blokov (v tomto prípade počtom betónových základových blokov). Týmto výpočtom získame trvanie každej úlohy. Potom tieto trvania vynesieme na časovú os, počnúc dátumom začiatku projektu, vytvoríme graf a výsledkom je vizuálne znázornenie, ktoré ukazuje, kedy by sa mala každá úloha začať a skončiť. Parameter "Poradie prác" pre procesy nám navyše umožňuje pochopiť, či pracovný proces prebieha paralelne ("Poradie prác" napr. 1.1-1.1) alebo sekvenčne (1.1-1.2).

**Gantsov diagram** je grafický nástroj na plánovanie a riadenie projektov, ktorý predstavuje úlohy ako vodorovné stĺpce na časovej osi. Každý stĺpec predstavuje trvanie úlohy, jej začiatok a koniec.

Harmonogram alebo Gantsov diagram pomáha projektovým manažérom a pracovníkom jasne pochopiť, kedy a v akom poradí sa majú jednotlivé fázy výstavby vykonávať, čím sa zabezpečí efektívne využívanie zdrojov a dodržiavanie termínov.

Uveďme si časový plán montáže troch betónových základových blokov s použitím výpočtov z vyššie uvedenej tabuľky. Na základe tabuľky nákladov (obr. 5.3-1) z vyššie uvedeného príkladu požiadajte

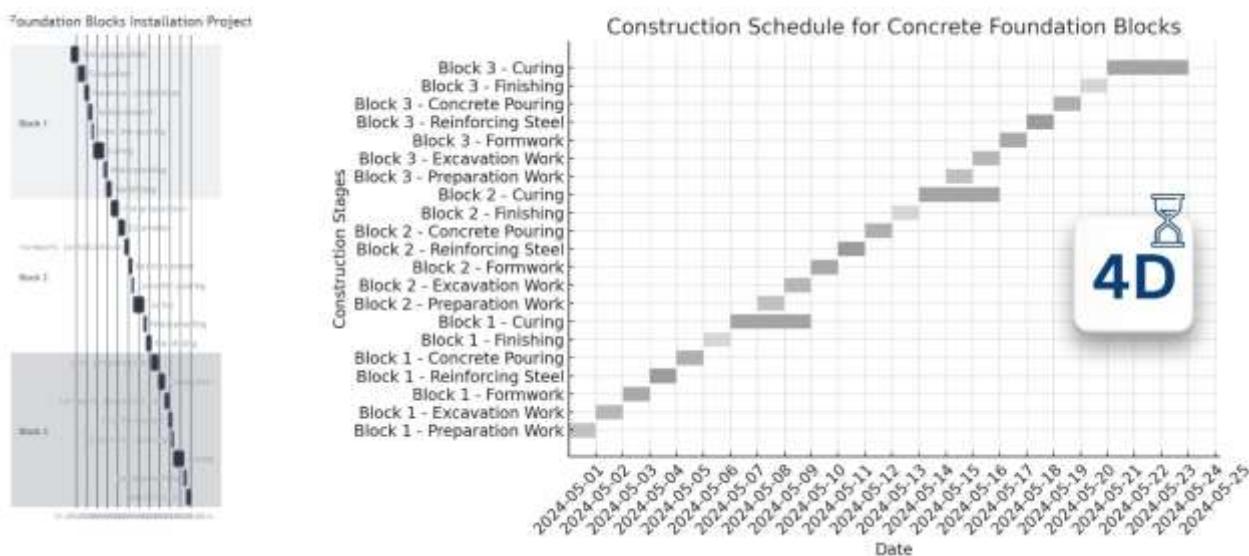
LLM, aby naplanoval inštaláciu 3 prvkov základových blokov, napr. na prvý máj 2024.

Ak chceme poslať kalkuláciu nákladov do LLM, môžeme nahrať tabuľku kalkulácie nákladov vo formáte XLSX alebo jednoducho vložiť snímku obrazovky vo formáte JPEG s kalkuláciou nákladov priamo do chatu LLM (obr. 5.3-1). LLM samostatne nájde knižnicu na vizualizáciu obrázku tabuľky a vynásobením časových atribútov činností z tabuľky ich objemom doplní všetky údaje do grafu.

- ❷ Pošlite textovú žiadosť na adresu LLM:

Vytvorte harmonogram prác vo forme Ganttovoho diagramu s použitím hodnôt z časovo súvisiacej tabuľky (obr. 5.3-1 ako JPEG v prílohe) pre inštaláciu 3 základových blokov. Práce pre každý blok sa vykonávajú postupne. Začiatok prác sa stanoví od 01.05.2024 ↵

- ❸ Odpoveď LLM



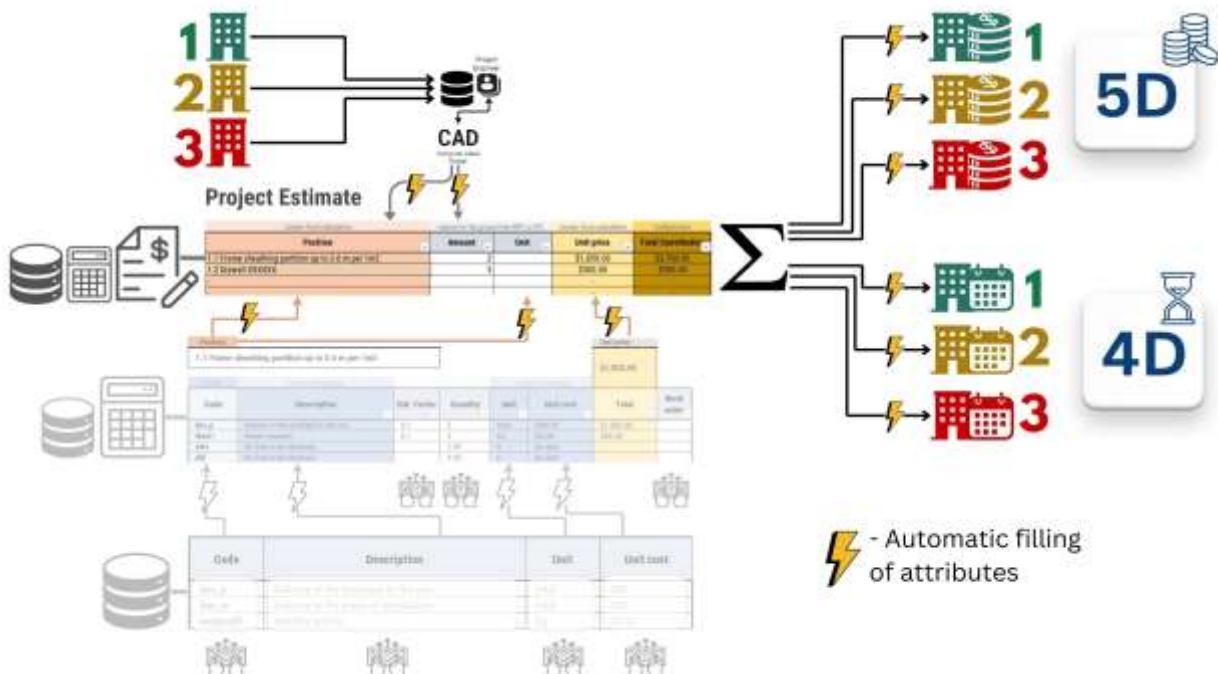
Obr. 5.3-2 Automaticky vygenerované viacerými LLM Ganttovo diagram zobrazuje etapy výstavby troch betónových blokov podľa podmienok z výzvy.

Výsledný graf (obr. 5.3-2) je časový diagram, v ktorom každý vodorovný stĺpec zodpovedá určitej fáze prác na základovom bloku a zobrazuje postupnosť operácií (parameter "Poradie prác"), ako sú príprava, výkop, montáž debnenia, armovanie, betonáž a dokončovacie práce,, t. j. tie procesy, ktoré majú vo výpočtoch vyplnené časové parametre a postupnosť.

Takýto harmonogram (obr. 5.3-2) nezohľadňuje obmedzenia súvisiace s pracovnými dňami, zmenami alebo normami pracovného času, ale je určený výlučne na koncepcnú vizualizáciu procesu. Presný harmonogram, ktorý bude odrážať súbežnosť prác, možno doplniť vhodnými podnetmi alebo dodatočnými pokynmi v rámci chatu

Pomocou jednotnej kalkulácie nákladov (obr. 5.3-1) je možné vďaka objemovým atribútom z 3D -

geometrie automaticky odhadnúť náklady projektu prostredníctvom automatizovaných odhadov a zároveň vypočítať časové charakteristiky skupín vo forme tabuľiek alebo grafov pre rôzne varianty projektu (obr. 5.3-3).



Obr. 5.3-3 Automatický výpočet, umožňuje okamžité a automatické predpovedanie nákladov a času pre rôzne možnosti projektu.

Moderné modulárne ERP -systémy (obr. 5.4-4), ktoré načítavajú údaje z modelov CAD, používajú podobné metódy automatizovaného výpočtu času, ktoré výrazne skracujú rozhodovací proces. To umožňuje okamžite a presne naplánovať harmonogramy prác a vypočítať celkový čas potrebný na splnenie všetkých úloh pri realizácii projektu so zohľadnením reálnych cien.

### Rozšírené vrstvy atribútov 6D -8D: od energetickej účinnosti po zaistenie bezpečnosti

6D, 7D a 8D sú rozšírené úrovne informačného modelovania, z ktorých každá prispieva ďalšími vrstvami atribútov do komplexného informačného modelu projektu, ktorého základom sú atribúty 3D - modelu s ich počtom a rozsahom. Každá ďalšia vrstva zavádzá špecifické parametre, ktoré sú potrebné na ďalšie zoskupovanie alebo ďalšiu identifikáciu v iných systémoch, ako sú napríklad systémy správy nehnuteľností (PMS), počítačom podporované riadenie zariadení (CAFM), riadenie stavebných projektov (CPM) a systémy riadenia bezpečnosti (SMS).



Obrázok 5.3-4 Atribúty 6D, 7D a 8D v informačnom modeli údajov rozširujú zohľadnenie rôznych aspektov projektu, od energetickej účinnosti až po bezpečnosť.

- **6D** sa okrem databázy projektov (alebo dátového rámca (obrázok 4.1-13)) s geometrickými a objemovými atribútmi prvkov pridávajú informácie (stĺpce atribútov) o environmentálnej udržateľnosti. Patria sem informácie týkajúce sa energetickej účinnosti, uhlíkovej stopy, recyklovateľnosti materiálov a používania technológií šetrných k životnému prostrediu. Tieto údaje umožňujú posúdiť vplyv projektu na životné prostredie, optimalizovať projektové rozhodnutia a dosiahnuť ciele trvalo udržateľného rozvoja (ESG).
- **7D** atribúty dopĺňajú atribúty potrebné na riadenie údržby budov. Ide o údaje o plánoch údržby, životných cykloch komponentov, technickej dokumentácii a histórii opráv. Tento súbor informácií zabezpečuje, že model možno integrovať so systémami údržby (CAFM, AMS), umožňuje efektívne plánovanie údržby, výmenu zariadení a poskytuje podporu počas celého životného cyklu objektu.
- **8D** ďalšia vrstva atribútov, - obsahuje informácie týkajúce sa bezpečnosti - vo fáze výstavby aj počas následnej prevádzky. Model zahŕňa opatrenia na zaistenie bezpečnosti personálu, pokyny pre prípad núdze, evakuačné systémy a požiadavky na požiarunu ochranu. Integrácia týchto údajov do digitálneho modelu pomáha vopred zohľadniť riziká a vypracovať architektonické, technické a organizačné riešenia, ktoré zohľadňujú požiadavky na bezpečnosť a ochranu zdravia.

V štruktúrovanej tabuľkovej forme predstavujú vrstvy 4D až 8D ďalšie atribúty vo forme stĺpcov s vyplnenými hodnotami (obrázok 5.3-5), ktoré sú pridané k už vyplneným atribútom 3D -modelu, ako sú názov, kategória, typ a objemové charakteristiky. Hodnoty vo vrstvách atribútov 6D, 7D a 8D obsahujú ďalšie textové a číselné údaje, ako je percento recyklácie, uhlíková stopa, záručná doba, cyklus výmeny, dátum inštalácie, bezpečnostné protokoly atď.

ID	Type Name	Width	Length	Recyclability	Carbon Footprint	Warranty Period	Replacement Cycle	Maintenance Schedule	Installation Date	Wellbeing Factors	Safety Protocols
W-NEW	Window	120 cm	-	90%	1622 kgCO <sub>2</sub> e	8 years	20 years	Annual	-mon	XYZ Windows	ISO 45001
W-OLD1	Window	100 cm	140 cm	90%	1522 kgCO <sub>2</sub> e	8 years	15 years	Biannual	08/22/2024	XYZ Windows	OSHA Standard
W-OLD2	Window	110 cm	160 cm	90%	1522 kgCO <sub>2</sub> e	-	15 years	Biannual	08/24/2024	????	OSHA Standard
D-122	Door	90 cm	210 cm	100%	1322 kgCO <sub>2</sub> e	15 years	25 years	Biennial	08/25/2024	Doors Ltd.	OSHA Standard

Obrázok 5.3-5 6D -8D pridať atribútové vrstvy do informačného modelu údajov, ktorý už obsahuje geometrické a objemové atribúty z 3D -modelu.

Pre naše nové okno (obrázok 4.4-1) môže mať prvok s identifikátorom W-NEW (obrázok 5.3-5) tieto atribúty 3D -8D:

**3D -atribúty - geometrické informácie získané zo systémov CAD:**

- "Názov typu" - prvok "Window"
- "Šírka" - 120 cm
- Okrem toho môžete pridať body "Bounding Box" prvku alebo jeho "geometrie BREP / MESH" ako samostatný atribút.

**Atribúty 6D - environmentálna udržateľnosť:**

- 90-percentná miera recyklácie
- "Uhlíková stopa - 1 622 kg CO<sub>2</sub>

**Atribúty 7D - údaje o správe objektov:**

- "Záručná doba" - 8 rokov
- "Náhradný cyklus" má 20 rokov
- "Údržba" - vyžaduje sa každoročne

**Atribúty 8D - zabezpečenie bezpečného užívania a prevádzky budov:**

- Okno "Nainštalované" - spoločnosť "XYZ Windows"
- "Bezpečnostný štandard" - v súlade s normou ISO 45001

Všetky parametre zaznamenané v databáze alebo súbore údajov (obr. 5.3-5) potrebujú odborníci v rôznych oddeleniach na zoskupovanie, vyhľadávanie alebo výpočty. Takýto viacrozumný atribútový opis objektov projektu poskytuje úplný obraz o ich životnom cykle, prevádzkových požiadavkách a mnohých ďalších aspektoch potrebných na projektovanie, výstavbu a prevádzku.

## Odhad emisií CO<sub>2</sub> a výpočet emisií oxidu uhličitého zo stavebných projektov

Popri téme udržateľnosti stavebných projektov v etape 6D (obrázok 5.3-5) sa moderná výstavba zameriava na environmentálnu udržateľnosť projektov, kde sa jedným z kľúčových aspektov stáva hodnotenie a minimalizácia emisií oxidu uhličitého CO<sub>2</sub>, ktoré vznikajú počas etáp životného cyklu projektu (napr. výroba a montáž).

Odhad a výpočet emisií uhlíka zo stavebných materiálov je proces, pri ktorom sa celkové emisie uhlíka určia vynásobením objemových vlastností prvku alebo skupiny prvkov použitých v projekte vhodným faktorom emisií uhlíka pre danú kategóriu.

Zohľadnenie emisií uhlíka pri hodnotení stavebných projektov ako súčasť širších kritérií ESG (environmentálne, sociálne a riadenie) pridáva do analýzy novú úroveň zložitosti. Je to dôležité najmä pre klienta - investora pri získavaní príslušných certifikátov, ako sú LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) alebo DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Získanie jednej z týchto certifikácií môže výrazne zvýšiť predajnosť nehnuteľnosti, zjednodušiť uvedenie do prevádzky a zabezpečiť súlad s požiadavkami nájomcov orientovaných na udržateľnosť (ESG). V závislosti od požiadaviek projektu možno použiť aj certifikáty HQE (Haute Qualité Environnementale, francúzsky štandard pre zelené budovy), WELL (WELL Building Standard, zameraný na zdravie a pohodlie užívateľov) a GRESB (Global Real Estate Sustainability Benchmark)

Environmentálne, sociálne a riadiace **aspekty ESG** (environmental, social and governance) je široký súbor zásad, ktoré možno použiť na interné aj externé hodnotenie podnikového riadenia, sociálneho a environmentálneho vplyvu podniku.

ESG, pôvodne vyvinutý začiatkom roku 2000 finančnými fondmi s cieľom poskytnúť investorom informácie o širokých environmentálnych, sociálnych a riadiacich kritériách, sa vyvinul v kľúčový ukazovateľ na hodnotenie spoločností aj projektov vrátane stavebných projektov. Podľa prieskumu významných poradenských firiem sa environmentálne, sociálne a riadiace aspekty (ESG) stávajú neoddeliteľnou súčasťou stavebného priemyslu.

Podľa EY (2023) "The Path to Carbon Neutrality" (Cesta k uhlíkovej neutralite) spoločnosti, ktoré aktívne implementujú princípy ESG, nielenže znížujú dlhodobé riziká, ale aj zvyšujú efektívnosť svojich obchodných modelov, čo je obzvlášť dôležité pri globálnej transformácii trhov [103]. V správe spoločnosti PwC o informovanosti o ESG sa uvádzá, že povedomie spoločností o dôležitosti ESG -faktorov sa pohybuje od 67% do 97%, pričom väčšina organizácií považuje tieto trendy za kľúčové pre budúcu udržateľnosť [104] a že podniky väčšinou zaznamenávajú výrazný tlak zo strany zainteresovaných strán na integráciu princípov ESG.

Integrácia princípov ESG do stavebných projektov tak prispieva nielen k získaniu medzinárodných certifikátov udržateľnosti, ako sú LEED, BREEAM, DGNB, ale zabezpečuje aj dlhodobú udržateľnosť a konkurencieschopnosť spoločností v tomto odvetví.

Jedným z najvýznamnejších faktorov, ktoré ovplyvňujú celkovú uhlíkovú stopu stavebného projektu, je výroba a logistika stavebných materiálov a komponentov. Materiály používané na stavbe majú často rozhodujúci vplyv na celkové emisie CO<sub>2</sub>, najmä v počiatočných fázach životného cyklu projektu - od ťažby surovín až po ich dodávku na stavenisko.

Výpočet emisií podľa kategórie alebo typu stavebného prvku si vyžaduje použitie referenčných emisných faktorov uhlíka, ktoré odrážajú množstvo CO<sub>2</sub> vyprodukovaného pri výrobe rôznych materiálov. Medzi takéto materiály patrí betón, tehly, recyklovaná oceľ, hliník a iné. Tieto hodnoty sa zvyčajne získavajú z renomovaných zdrojov a medzinárodných databáz, ako sú napríklad UK ICE 2015 (Inventory of Carbon and Energy) a US EPA 2006 (U.S. Environmental Protection Agency) [105]. V nasledujúcej tabuľke (obrázok 5.3-6) sú zhrnuté základné emisné faktory pre celý rad bežných stavebných materiálov. Pre každý materiál sú uvedené dva kľúčové parametre: špecifické emisie CO<sub>2</sub> (v kilogramoch na kilogram materiálu) a konverzné faktory objemu na hmotnosť (v kilogramoch na meter kubický), ktoré sú potrebné na integráciu výpočtov do modelu návrhu a prepojenie so zoskupením údajov QTO.



**CO<sub>2</sub>**

<b>Carbon Emitted in Production</b>		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	<b>Coefficient m<sup>3</sup> to kg</b>
<b>Material</b>	<b>Abbreviated</b>	Process Emissions (kg CO <sub>2</sub> e/ kg of product) (K1)	Process Emissions (kg CO <sub>2</sub> e/ kg of product) (K2)	Kg / m <sup>3</sup> (K3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

**Obrázok 5.3-6 Množstvo uhlíka emitovaného pri výrobe rôznych stavebných materiálov podľa databázy ICE Spojeného kráľovstva a US EPA.**

Na výpočet celkových emisií CO<sub>2</sub> pre projekt je rovnako ako pri výpočtoch 4D a 5D potrebné určiť atribútové objemy každej skupiny zariadení. To možno vykonať pomocou nástrojov kvantitatívnej analýzy (QTO) získaním atribútových objemov v metroch kubických, ako sa podrobne uvádzajú v časti o odoberaní množstva. Tieto objemy sa potom vynásobia príslušnými koeficientmi pre atribút "emisie CO<sub>2</sub> z procesov" každej skupiny materiálov.

- ⌚ Automaticky extrahujme tabuľku objemov podľa typu prvkov z projektu CAD (BIM) zoskupením všetkých údajov projektu, ako to už bolo urobené v predchádzajúcich kapitolách. Na vykonanie tejto úlohy sa obráťme na LLM.

Zoskupte, prosím, tabuľku DataFrame z projektu CAD (BIM) podľa parametra stĺpca "Názov objektu" (alebo "Typ") a zobrazte počet položiek v každej skupine a zhrňte parameter "Objem" pre všetky položky v danom type. ↵

💡 Odpoveď LLM :

```

Create in LLM Chat
Data
f(x)
Output
Run in IDE
  
```

```

1 import pandas as pd
2
3 # Import Projekt as CSV tain DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg(Amount='size', Sum of Volumes, m³='sum')
  
```

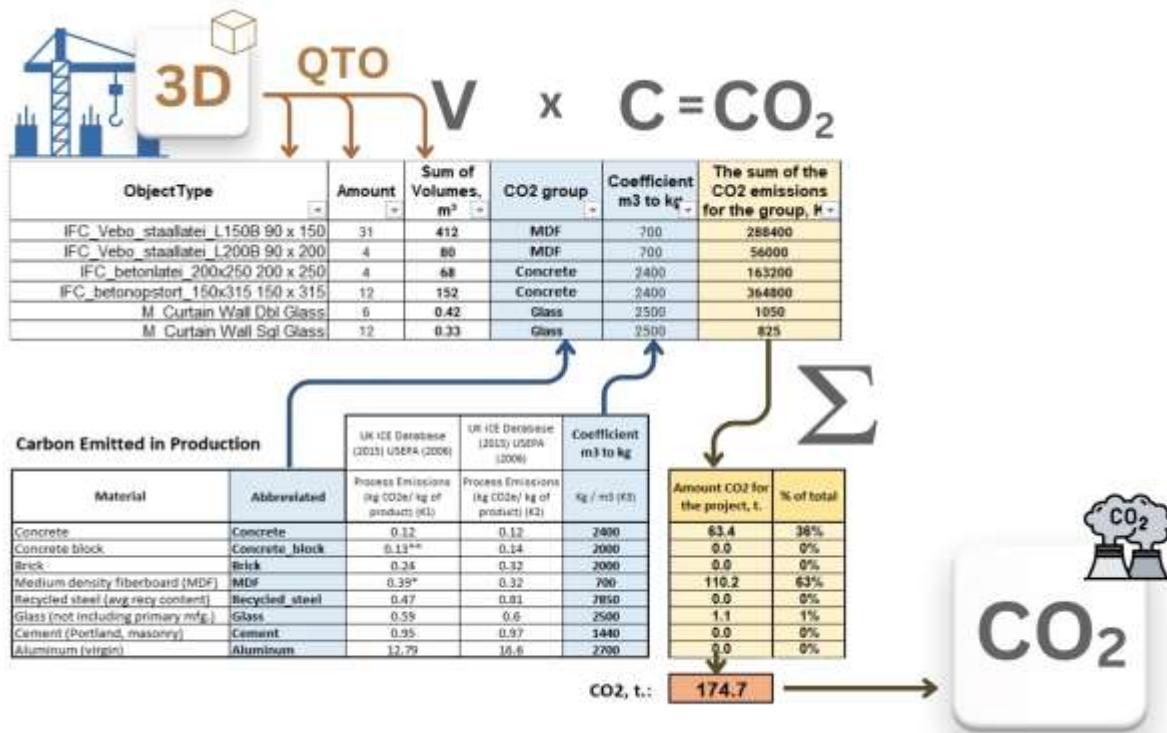
ObjectType	Amount	Sum of Volumes, m <sup>3</sup>
IFC_Vebo_stallatei_L150B 90 x 150	31	412
IFC_Vebo_stallatei_L200B 90 x 200	4	80
IFC_betonlatei_200x250 200 x 250	4	68
IFC_betonopstort_150x315 150 x 315	12	152
M_Curtain Wall Dbl Glass	6	0.42
M_Curtain Wall Sgl Glass	12	0.33

Obrázok 5.3-7 Vygenerovaný kód v LLM nám zoskupil entity projektu podľa typu (ObjectType) so súhrnným atribútom "Volume".

Ak chcete automatizovať výpočet celkových emisií CO<sub>2</sub> pre celý projekt, stačí nastaviť automatické mapovanie údajov v tabuľke alebo manuálne priradiť typy prvkov (obrázok 5.3-7) k príslušným typom materiálov (obrázok 5.3-6) z tabuľky emisných faktorov. Hotovú tabuľku s emisnými faktormi a vzorcami, ako aj kód na načítanie objemov z formátov CAD (BIM) a automatizované určovanie CO<sub>2</sub>, nájdete na GitHube vyhľadaním "CO<sub>2</sub>\_calculating-the-embodied-carbon. DataDrivenConstruction." [106].

Integrácia údajov po zoskupení prvkov QTO z databázy CAD tak umožňuje automatický výpočet emisií oxidu uhličitého (obr. 5.3-8) pre rôzne varianty návrhu. To umožňuje analyzovať vplyv rôznych materiálov v rôznych variantoch a vybrať len tie riešenia, ktoré spĺňajú požiadavky zákazníka na emisie CO<sub>2</sub> na získanie konkrétneho certifikátu pri uvedení budovy do prevádzky

Odhad emisií CO<sub>2</sub> vynásobením faktorov objemom zoskupených prvkov projektu je typickým príkladom úlohy v procese získavania hodnotenia ESG stavebnou spoločnosťou (napr. certifikácia LEED) pre objekt.



Obrázok 5.3-8 Integrácia skupín QTO z databáz CAD poskytuje presnosť a automatizáciu pri odvodzovaní odhadov konečných emisií CO<sub>2</sub>.

Podobne, definovaním objemov skupín prvkov môžeme vykonávať výpočty pre kontrolu a logistiku materiálu, monitorovanie a riadenie kvality, energetické modelovanie a analýzu a množstvo ďalších úloh na získanie nového stavu atribútu (parametra v tabuľke) pre jednotlivé skupiny prvkov aj pre celý projekt.

Ak počet takýchto výpočtových procesov v podniku začne narastať, vzniká otázka potreby automatizácie takýchto výpočtov a implementácie výsledkov výpočtov do podnikových procesov a systémov správy údajov.

Vzhľadom na zložitosť kompletného riešenia zadávajú stredné a veľké spoločnosti v stavebnictve takúto automatizáciu spoločnostiam využívajúcim systémy ERP (alebo PMIS). Vývojové spoločnosti vytvárajú pre veľkých klientov jeden komplexný modulárny systém na správu mnohých rôznych informačných vrstiev vrátane kalkulácií materiálov a zdrojov.



## KAPITOLA 5.4.

# STAVEBNÝ ERP A PMIS SYSTÉMY

### Stavebný ERP -systémy na príklade výpočtov a odhadov

Modulárne systémy ERP integrujú rôzne atribútové (informačné) vrstvy a dátové toky do jedného komplexného systému, ktorý umožňuje projektovým manažérom synchronizované riadiť zdroje, finančie, logistiku a ďalšie aspekty projektu v rámci jednej platformy. Stavebný systém ERP pôsobí ako "mozog" stavebných projektov, zjednodušuje opakujúce sa procesy prostredníctvom automatizácie, poskytuje transparentnosť a kontrolu v celom procese výstavby.

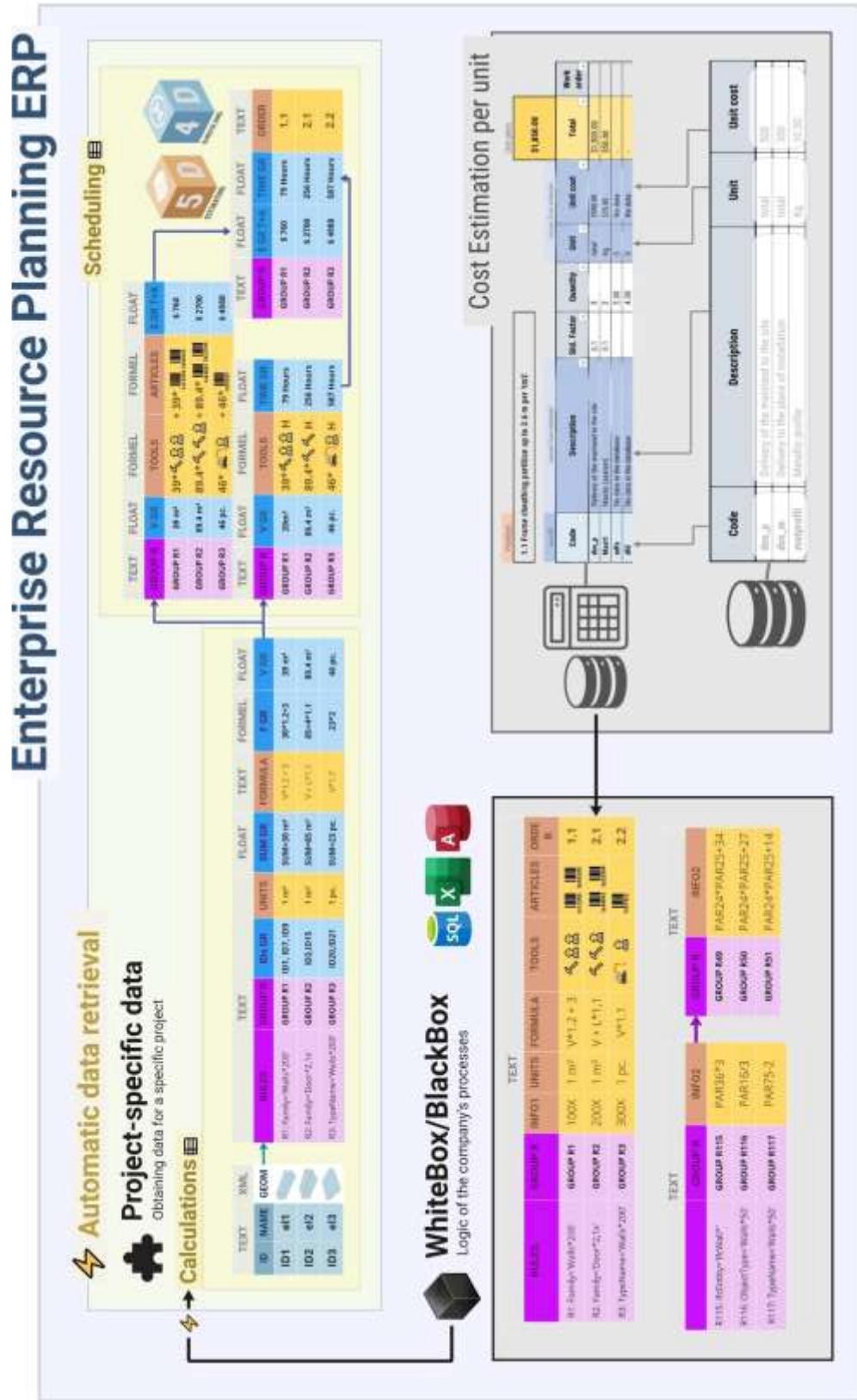
**Stavebné ERP -systémy** (Enterprise Resource Planning) sú komplexné softvérové riešenia určené na riadenie a optimalizáciu rôznych aspektov stavebného procesu. Jadrom stavebných systémov ERP sú moduly na riadenie kalkulácie nákladov a plánovania, ktoré z nich robia dôležitý nástroj na efektívne plánovanie zdrojov.

Moduly ERP -systémy umožňujú používateľom zadávať, spracovávať a analyzovať údaje štruktúrovaným spôsobom, ktoré pokrývajú rôzne aspekty projektu, ktoré môžu zahrňať evidenciu materiálových a mzdových nákladov, využitie zariadení, riadenie logistiky, ľudské zdroje, kontakty a ďalšie stavebné činnosti.

Jedným z funkčných blokov systému je modul automatizácie obchodnej logiky - BlackBox/WhiteBox, ktorý plní úlohu centra riadenia procesov.

BlackBox /WhiteBox umožňuje špecialistom používajúcim ERP -systém flexibilne spravovať prostredníctvom prístupových práv rôzne aspekty podnikania, ktoré už boli vopred nakonfigurované inými používateľmi alebo správcami. V kontexte systémov ERP sa pojmy *BlackBox* a *WhiteBox* vzťahujú na úroveň transparentnosti a kontrolovateľnosti vnútornej logiky systému:

- **BlackBox** ("čierna skrinka") - používateľ komunikuje so systémom prostredníctvom rozhrania bez prístupu k vnútornej logike vykonávania procesu. Systém vykonáva výpočty sám na základe vopred definovaných pravidiel, ktoré sú pred koncovým používateľom skryté. Zadá údaje a dostane výsledok bez toho, aby vedel, aké atribúty alebo koeficienty boli použité vo vnútri.
- **WhiteBox** ("biela skrinka") - logika procesu je k dispozícii na zobrazenie, prispôsobenie a úpravu. Pokročilí používatelia, správcovia alebo integrátori môžu ručne definovať algoritmy spracovania údajov, pravidlá výpočtu a scenáre interakcie medzi subjektmi projektu.



Obr. 5.4-1 Architektúra stavebného ERP -systému, na získanie odhadov a harmonogramov prác ručným vyplnením atribútov objemu.

Príkladom je prípad, keď skúsený používateľ alebo správca nastaví pravidlo: ktoré atribúty v odhadе sa majú vynásobiť medzi sebou alebo zoskupiť podľa určitej vlastnosti a kde sa má zaznamenať konečný výsledok. Neskôr menej vyškolení odborníci, napríklad odhadcovia, jednoducho načítajú nové údaje do ERP prostredníctvom používateľského rozhrania - a získajú hotové odhady, harmonogramy alebo špecifikácie bez toho, aby museli písanie kód alebo rozumieť technickým detailom logiky.

V predchádzajúcich kapitolách sa v súvislosti s interakciou LLM hovorilo o výpočtových a logických moduloch. V prostredí ERP sa takéto výpočty a transformácie uskutočňujú v rámci modulov skrytých za rozhraním tlačidiel a formulárov.

V nasledujúcom príklade (obr. 5.4-1) správca ERP -systému v module BlackBox /WhiteBox definoval pravidlá pre párovanie atribútov entít z odhadov s atribútmi pre zoskupenie QTO. Vďaka tomuto nastaveniu modulu BlackBox/WhiteBox (manažérom alebo správcom) používateľ (odhadca alebo inžinier) manuálnym pridaním atribútu množstva alebo objemu prostredníctvom používateľského rozhrania ERP automaticky získá hotové odhady a harmonogramy prác. Týmto spôsobom sa procesy výpočtu a generovania odhadov, o ktorých sa hovorilo v predchádzajúcich kapitolách, pomocou kódu, vnútri ERP, stávajú poloautomatickým dopravníkom.

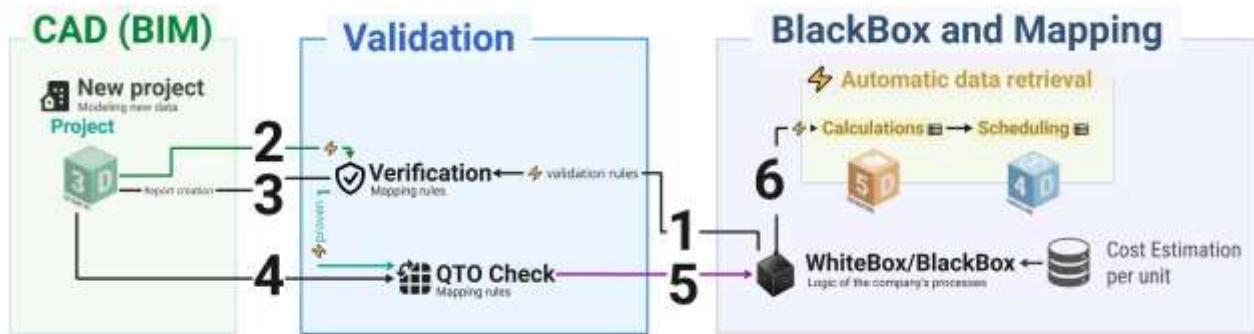
Pripojením tohto poloautomatizovaného procesu k objemovým atribútom z modelov CAD (BIM) (obr. 4.1-13), napríklad prostredníctvom načítania projektu CAD do modulu ERP, ktorý je na tento účel vopred nakonfigurovaný, sa tok údajov zmení na synchronizovaný mechanizmus schopný samostatne a okamžite aktualizovať hodnoty jednotlivých skupín prvkov alebo celého projektu v reakcii na akékoľvek zmeny v ňom počas fázy projektovania pri načítaní modelu CAD do ERP.

Aby bolo možné vytvoriť automatizovaný tok údajov (obr. 5.4-2) medzi systémami CAD (BIM) a ERP, je potrebné štruktúrovane definovať základné procesy a požiadavky na údaje z databáz modelov CAD (BIM), ako sme už uviedli v predchádzajúcej kapitole "Požiadavky a zabezpečenie kvality údajov ". Tento proces v ERP je rozdelený do podobných krokov:

- **Vytvorenie validačných pravidiel (1)**, ktoré zohrávajú dôležitú úlohu pri zabezpečovaní správnosti údajov vstupujúcich do systému ERP. Validačné pravidlá slúžia ako filtre, ktoré overujú entity a ich atribúty a umožňujú vstup do systému len tým položkám, ktoré spĺňajú požiadavky. Viac informácií o overovaní a validácii sa dozviete v kapitole "Vytváranie požiadaviek a validácia kvality údajov ".
- Potom sa v ERP uskutoční **proces overovania (2)**, ktorý potvrdí, že všetky prvky projektových entít s ich atribútmi a hodnotami boli vytvorené správne a sú pripravené na ďalšie kroky spracovania.
- Ak sa vyskytnú problémy s neúplnými údajmi atribútov, **vytvorí sa správa (3)** a projekt sa spolu s pokynmi na opravu odošle na revíziu, kym nebude pripravený na ďalšiu iteráciu.
- Po validácii a overení údajov o projekte sa tieto údaje použijú v ďalšom module ERP **(4) na vytvorenie tabuľiek preberaného množstva (QTO)**, ktoré vytvárajú atribúty množstva pre skupiny subjektov, materiály a zdroje podľa vopred vytvorených pravidiel (WhiteBox/BlackBox).
- Zoskupené údaje podľa pravidiel párovania alebo QTO sú automaticky **integrované s**

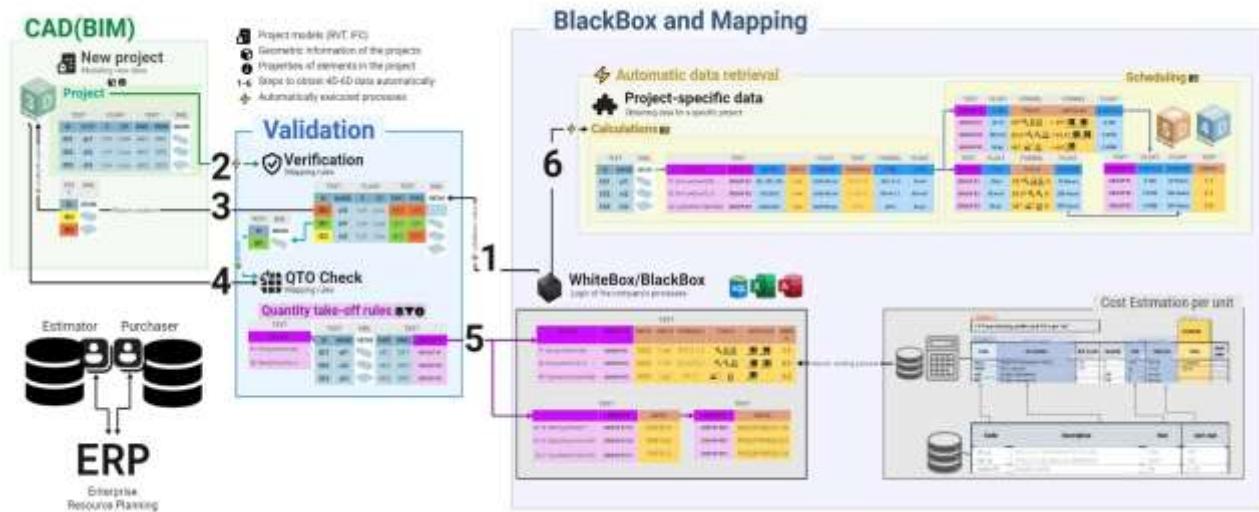
výpočtami (napr. náklady a čas) (5).

- V poslednom kroku ERP -systému používateľ vynásobením atribútov rozsahu z tabuľky QTO s atribútm procesných tabuliek (napr. odhadované položky), **automaticky vygeneruje výsledky výpočtov** (6) (napr. odhady nákladov, harmonogramy prác alebo emisie CO<sub>2</sub>) pre každú skupinu subjektov a pre projekt ako celok.

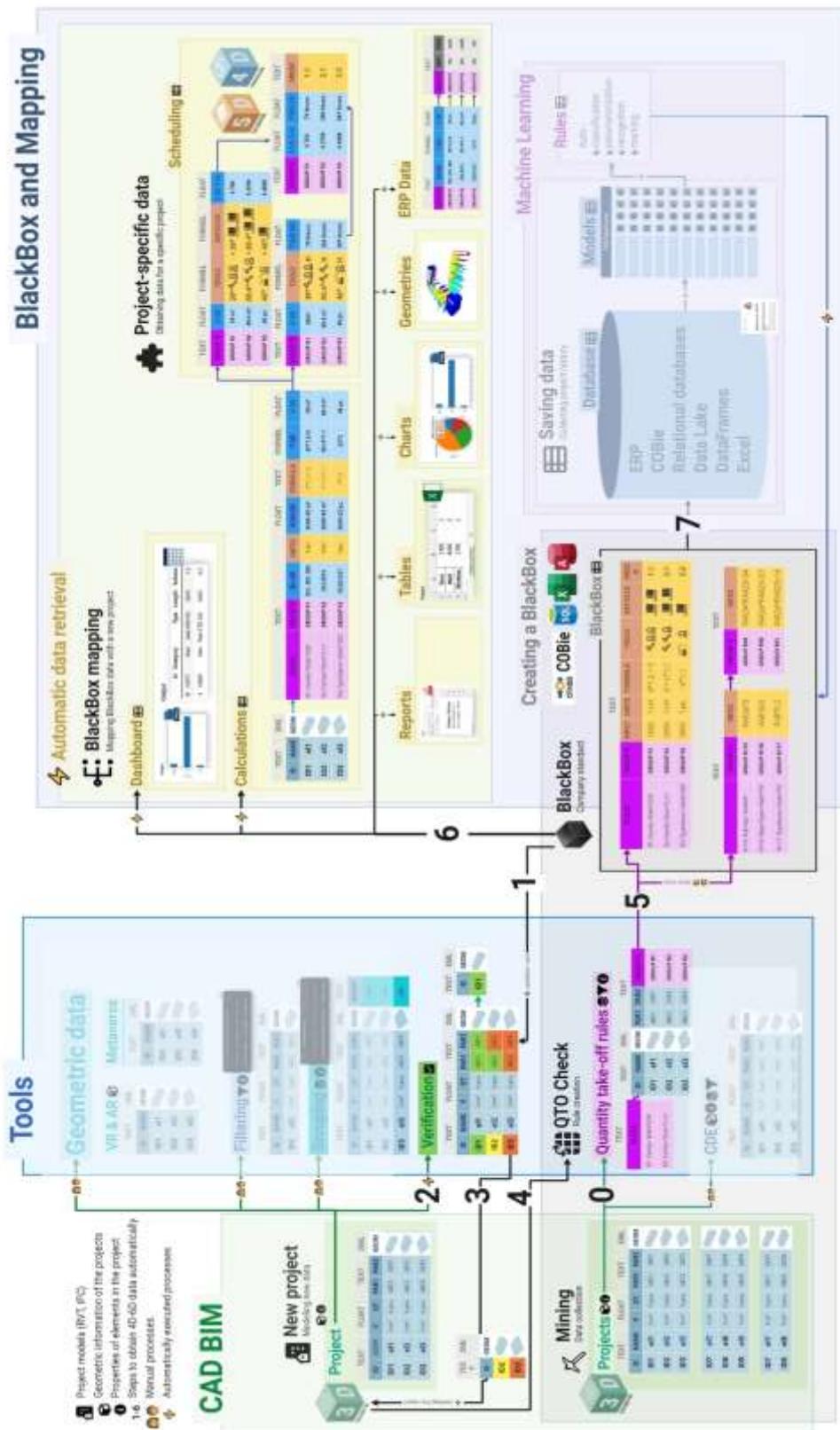


Obr. 5.4-2 Architektúra stavebného ERP -systému s CAD (BIM), od vytvorenia pravidiel validácie (1) až po automatický výpočet nákladov a harmonogramov prác (5-6).

V modulárnom systéme ERP sú procesy integrované pomocou softvéru, ktorý obsahuje používateľské rozhranie. Za rozhraním sa nachádza back-end, kde štruktúrované tabuľky spracúvajú údaje vykonávaním rôznych operácií, ktoré boli vopred nakonfigurované manažérom alebo správcom. Výsledkom je, že používateľ vďaka preddefinovanej a prispôsobenej logike automatizácie (v moduloch BlackBox /WhiteBox) dostáva poloautomaticky pripravené dokumenty, ktoré spĺňajú jeho úlohy



Obr. 5.4-3 ERP -systém pomáha manažérom a používateľom pohybovať sa medzi odbornými tabuľkami a vytvárať nové údaje.



Obr. 5.4-4 ERP -systém je integrovaný s analytickými nástrojmi a automatizuje rozhodovací proces spoločnosti.

Podobne aj procesy v ERP -systémoch, od vzniku až po konečný výpočet (kroky 1-6 Obrázok 5.4-3), predstavujú reťazec vzájomne prepojených krovov, ktoré v konečnom dôsledku zabezpečujú transparentnosť, efektívnosť a presnosť plánovania.

Moderné stavebné ERP -systémy obsahujú nie len moduly na výpočet nákladov a harmonogramu, ale aj desiatky ďalších predkonfigurovaných modulov, ktoré zvyčajne pokrývajú správu dokumentov, sledovanie priebehu projektu, riadenie zmlúv, dodávateľský reťazec a logistiku, ako aj integráciu s inými podnikovými systémami a platformami. Integrované analytické nástroje ERP umožňujú používateľom automatizovať tvorbu informačných panelov na monitorovanie projektových KPI (KPI - klúčové ukazovatele výkonnosti). Tým sa zabezpečuje centralizované a konzistentné riadenie všetkých aspektov stavebného projektu so snahou spojiť veľké množstvo aplikácií a systémov na jednej platforme.

V budúcnosti sa bude ERP -analytika používať v kombinácii so strojovým učením na zvýšenie presnosti a optimalizáciu procesu výpočtu budúcich atribútov projektu. Údaje a atribúty analyzované a zhromažďované zo systémov ERP v rámci Big Data (obr. 5.4-4) budú v budúcnosti tvoriť základ pre vytváranie prediktívnych modelov, ktoré dokážu presne predvídať potenciálne oneskorenia, riziká alebo napríklad možné zmeny v nákladoch na materiál.

Ako alternatíva k ERP, sa v stavebníctve často používa PMIS (Project Management Information System), systém riadenia projektov určený na podrobné riadenie úloh na úrovni jednotlivých stavebných projektov.

## **PMIS: sprostredkovateľ medzi ERP a stavbou**

Na rozdiel od ERP, ktorý pokrýva celý reťazec obchodných procesov spoločnosti, PMIS sa zameriava na riadenie konkrétnego projektu, monitorovanie časového harmonogramu, rozpočtov, zdrojov a dokumentácie.

PMIS (Project Management Information System) je softvér na riadenie stavebných projektov určený na plánovanie, sledovanie, analýzu a podávanie správ o všetkých aspektoch projektu.

PMIS umožňuje spravovať dokumenty, harmonogramy, rozpočty a na prvý pohľad sa môže zdať, že PMIS je duplicitným riešením k ERP, ale klúčový rozdiel je v úrovni riadenia:

- **ERP** sa zameriava na obchodné procesy spoločnosti ako celku: náklady, zmluvy, obstarávanie, ľudské zdroje a riadenie zdrojov na podnikovej úrovni.
- **PMIS** sa zameriava na riadenie jednotlivých projektov, zabezpečuje podrobné plánovanie, kontrolu zmien, podávanie správ a koordináciu účastníkov.

V mnohých prípadoch sú to práve ERP -systémy, ktoré už majú dostatočnú funkcionality, a implementácia PMIS sa stáva skôr otázkou pohodlia a preferencií spoločnosti. Mnohí dodávateľia a zákazníci nepoužívajú PMIS preto, že je to nevyhnutné, ale preto, že im to vnučuje dodávateľ alebo veľký zákazník, ktorý chce agregovať údaje na konkrénej platforme.

Je potrebné spomenúť, že v medzinárodnej terminológii pre riadenie stavebných projektov existujú ďalšie samostatné populárne pojmy, ako napríklad PLM (Product Lifecycle Management) a EPC a EPM (Engineering, Procurement and Construction Management) - kontraktačné metódy v stavebníctve.

Ak spoločnosť už používa ERP s modulmi projektového riadenia, zavedenie PMIS môže byť zbytočným prepojením, ktoré duplikuje funkcie. Ak však procesy nie sú automatizované a údaje sú roztriedené, PMIS môže byť vhodnejším a jednoduchším nástrojom na údržbu.

## **Špekulácie, zisk, izolovanosť a nedostatok transparentnosti v ERP a PMIS**

Napriek vonkajšej jednoduchosti rozhraní a postupov sú stavebné ERP a PMIS -systémy vo väčšine prípadov uzavreté a nepružné riešenia. Takéto systémy sa zvyčajne dodávajú ako predkonfigurovaný softvérový balík od jedného dodávateľa s obmedzeným prístupom k interným databázam a logike procesov.

Vývoj a riadenie takýchto systémov čoraz viac preberajú dodávateľia CAD-(BIM-), pretože ich databázy obsahujú informácie, ktoré vyžadujú systémy ERP: kvantitatívne a objemové atribúty prvkov projektu. Namiesto poskytovania prístupu k týmto údajom v otvorenom alebo strojovo čitateľnom formáte však dodávateľia ponúkajú len obmedzené používateľské scenáre a uzavretú logiku spracovania - preddefinovanú v rámci modulov BlackBox. To znižuje flexibilitu systému a bráni jeho prispôsobeniu špecifickým podmienkam projektu.

Obmedzená transparentnosť údajov zostáva jednou z hlavných výziev digitálnych procesov v stavebníctve. Uzavretá databázová architektúra, nedostatočný prístup k úplným súborom atribútov stavebných prvkov, zameranie na *moduly* automatizácie BlackBox a nedostatok otvorených rozhraní výrazne zvyšujú riziká byrokracie dokumentov. Takéto obmedzenia vytvárajú úzke miesta v rozhodovacom procese, stážajú overovanie informácií a otvárajú priestor na skrývanie údajov alebo špekulácie v rámci systémov ERP/PMIS. Používateľia zvyčajne získavajú len obmedzený prístup - či už ide o osekané rozhranie alebo čiastočné API - bez možnosti priamej interakcie s primárnymi zdrojmi údajov. To je obzvlášť dôležité, keď ide o parametre automaticky generované z projektov CAD, ako sú objemy, plochy a množstvá používané na výpočty QTO.

V dôsledku toho sa mnohé stavebné spoločnosti namiesto hľadania efektívnosti prostredníctvom automatizácie procesov, otvorených údajov, znižovania transakčných nákladov a vytvárania nových obchodných modelov zameriavajú na riadenie externých parametrov - manipuláciu s faktormi, korekčnými faktormi a metódami výpočtu ovplyvňujúcimi náklady na projekt v uzavretých platformách ERP/PMIS. To vytvára priestor pre špekulácie, skresľuje skutočné výrobné náklady a znižuje dôveru medzi všetkými účastníkmi stavebného procesu

V stavebníctve sa zisk tvorí ako rozdiel medzi výnosmi z dokončeného projektu a variabilnými nákladmi, ktoré zahŕňajú projektové, materiálové, mzdové a iné priame náklady priamo súvisiace s realizáciou projektu. Kľúčovým faktorom ovplyvňujúcim hodnotu týchto nákladov však nie je len technológia alebo logistika, ale aj rýchlosť a presnosť kalkulácií a kvalita manažérskych rozhodnutí v podniku.

Problém zhoršuje skutočnosť, že vo väčšine stavebných spoločností zostávajú procesy výpočtu nákladov netransparentné nielen pre zákazníkov, ale aj pre samotných zamestnancov, ktorí nie sú súčasťou oddelenia odhadu alebo finančného oddelenia. Takáto uzavretosť podporuje v rámci spoločnosti vytvorenie privilegovanej skupiny špecialistov - nositeľov "finančnej expertízy", ktorí majú výhradné právo upravovať atribúty a korekčné faktory v ERP/PMIS -systémoch. Títo zamestnanci spolu s vedúcimi podnikov môžu v skutočnosti kontrolovať finančnú logiku projektu.

Odhadcovia sa v takýchto podmienkach menia na "finančných žonglérov", ktorí balansujú medzi maximalizáciou zisku spoločnosti a potrebou zachovať konkurencieschopnú cenu pre klienta. Zároveň sa musia vyhýbať očividným a hrubým manipuláciám, aby nepoškodili dobré meno spoločnosti. Práve v tejto fáze sa stanovujú koeficienty, ktoré majú zakryť nadhodnotené objemy alebo náklady na materiál a práce.

V dôsledku toho hlavnou schémou zvyšovania efektívnosti a ziskovosti podnikov pôsobiacich v stavebníctve nie je automatizácia a zrýchlenie rozhodovacích procesov, ale špekulácie s cenami materiálov a prác (obr. 5.4-5). Nadhodnocovanie nákladov na práce a materiály sa uskutočňuje "šedým" účtovníctvom v uzavretých ERP /PMIS - systémoch nadhodnotením percent nad priemerné trhové ceny materiálov alebo objemov prác pomocou koeficientov (obr. 5.1-6), o ktorých sa hovorilo v kapitole "Zostavenie kalkulácií a výpočet nákladov na práce na základe zdrojovej základne".

Výsledkom je, že zákazník dostane kalkuláciu, ktorá neodráža skutočné náklady alebo rozsah prác, ale je derivátom mnohých skrytých interných koeficientov. Zároveň sú subdodávatelia v snahe splniť podhodnotené sadzby stanovené generálnym dodávateľom často nútení nakupovať lacnejšie a nekvalitné materiály, čo zhoršuje výslednú kvalitu stavby.

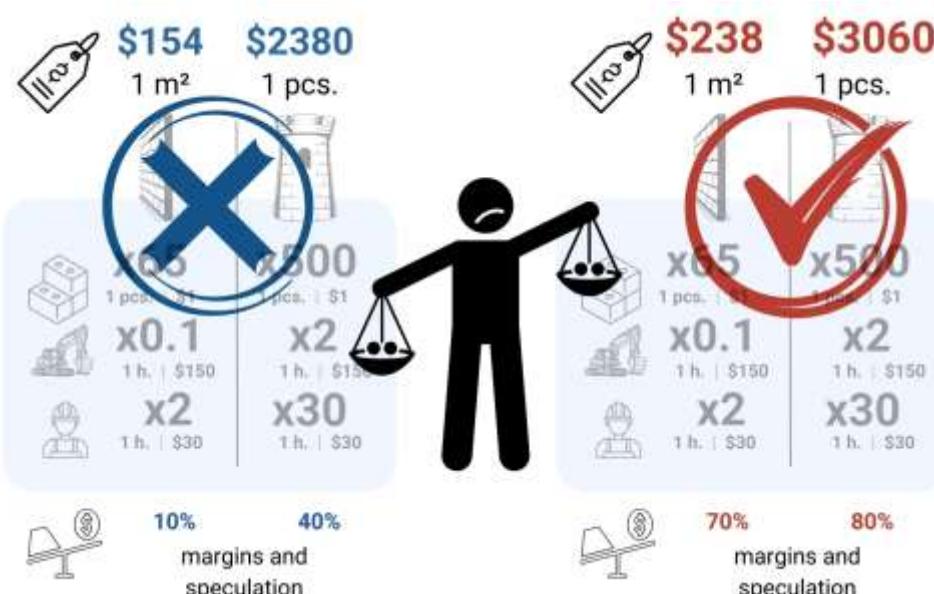
Špekulatívny proces hľadania ziskov zo vzduchu nakoniec poškodzuje klientov, ktorí dostávajú nespoľahlivé údaje, aj výkonných pracovníkov, ktorí sú nútenci hľadať ďalšie a ďalšie špekulatívne modely

Čím väčší je projekt, tým vyššia je úroveň byrokracie pri riadení údajov a procesov. Každý krok a každý modul často skrýva neprehľadné koeficienty a príplatky zakotvené vo výpočtových algoritnoch a interných postupoch. To nielenže sťažuje audit, ale aj výrazne skresľuje finančný obraz projektu. Vo veľkých stavebných projektoch takéto postupy často vedú k niekoľkonásobnému (niekedy až desaťnásobnému) zvýšeniu konečných nákladov, pričom skutočné objemy a náklady zostávajú mimo efektívnej kontroly klienta (obr. 2.1-3 Porovnanie plánovaných a skutočných nákladov veľkých

infraštrukturých projektov v Nemecku.

Podľa správy spoločnosti McKinsey & Company *Imagining the Digital Future of Construction* (2016) sa veľké stavebné projekty dokončujú v priemere o 20% neskôr, ako sa plánovalo, a až o 80% nad rámec rozpočtu [107].

Oddelenia odhadov a rozpočtov sa stávajú najstráženejším článkom v rámci spoločnosti. Prístup k nim je prísne obmedzený aj pre interných odborníkov a vzhľadom na uzavretú logiku a databázové štruktúry nie je možné objektívne posúdiť efektívnosť projektových rozhodnutí bez skreslenia. Netransparentnosť viedie k tomu, že podniky sú nútené neoptimalizovať procesy, ale bojovať o prežitie "kreatívnym" riadením čísel a faktorov (obr. 5.3-1, obr. 5.1-6 - napr. parameter "Bid. faktor").



Obr. 5.4-5 Špekulatívne pomery na úrovni osadenstva sú hlavným ziskom firiem a umením žonglovať medzi kvalitou práce a reputáciou.

To všetko spochybňuje ďalšie používanie uzavretých systémov ERP/PMIS v stavebníctve. V kontexte digitálnej transformácie a rastúcich požiadaviek na transparentnosť zo strany zákazníkov (obr. 10.2-3) je nepravdepodobné, že realizácia projektov v dlhodobom horizonte zostane závislá od proprietárnych riešení, ktoré obmedzujú flexibilitu, bránia integrácii a brzdia rozvoj podnikania.

A bez ohľadu na to, ako výhodné môže byť pre stavebné spoločnosti pracovať s dátovými silami a neprehľadnými údajmi v uzavretých databázach - budúcnosť stavebníctva bude nevyhnutne zahŕňať prechod na otvorené platformy, strojovo čitateľné a transparentné dátové štruktúry a automatizáciu založenú na dôvere. Táto transformácia bude riadená zhora - pod tlakom zákazníkov, regulačných orgánov a spoločnosti, ktorí čoraz viac požadujú zodpovednosť, udržateľnosť, transparentnosť a ekonomickú realizovateľnosť.

## Koniec éry uzavretých ERP /PMIS: stavebníctvo potrebuje nové prístupy

Používanie rozsiahlych modulárnych systémov ERP/PMIS -systémov pozostávajúcich z desiatok miliónov riadkov kódu mimoriadne sťaže akékoľvek zmeny v nich. V tomto prípade sa prechod na novú platformu za prítomnosti modulov už predkonfigurovaných pre podnik, desiatok tisíc článkov v databázach zdrojov (obr. 5.1-3) a tisícov hotových kalkulácií (obr. 5.1-6) mení na nákladný a zdĺhavý proces. Čím viac kódu a staršej architektúry - tým vyššia úroveň internej neefektívnosti a každý nový projekt situáciu len zhorší. V mnohých spoločnostiach sa migrácia údajov a integrácia nových riešení stáva niekoľkoročnou epopejou sprevádzanou neustálym prepracovávaním a nekonečným hľadaním kompromisov. Výsledkom je často návrat k starým, známym platformám napriek ich obmedzeniam.

Ako sa zdôrazňuje v nemeckej správe Black Book [108] o systémových chybách v riadení údajov v stavebníctve, roztrieštenosť informácií a chýbajúci centralizovaný prístup k ich riadeniu sú hlavnou príčinou neefektívnosti. Bez štandardizácie a integrácie strácajú údaje svoju hodnotu a stávajú sa skôr archívom než nástrojom riadenia.

Hlavnou príčinou straty kvality údajov je nedostatočné plánovanie a kontrola stavebných projektov, čo často vedie k výraznému zvýšeniu nákladov. V časti "Zaostrené na: Explózia nákladov" knihy Black Book sa analyzujú kľúčové faktory, ktoré prispievajú k týmto nežiaducim dôsledkom. Patria k nim nedostatočné analýzy potrieb, chýbajúce štúdie uskutočniteľnosti a nekoordinované plánovanie, ktoré vedú k dodatočným nákladom, ktorým sa dalo predísť.

Vo vyspelom ekosystéme IT spoločnosti je výmena zastaraného systému porovnatelná s výmenou nosného stípa v už postavenej budove. Nestačí len odstrániť starý a nainštalovať nový - je dôležité urobiť to tak, aby budova zostala stabilná, stropy sa nezrútili a všetky komunikácie naďalej fungovali. Práve v tom spočíva problém: každá chyba môže mať vážne dôsledky pre celý systém spoločnosti.

Napriek tomu vývojári veľkých produktov ERP pre stavebníctvo naďalej používajú množstvo napísaného kódu ako argument v prospech svojej platformy. Na špecializovaných konferenciách možno stále počuť vety typu: "Vytvorenie takéhoto systému by trvalo 150 človekorokov", a to napriek tomu, že väčšina funkcií takýchto systémov v sebe skrýva databázy a pomerne jednoduché funkcie na prácu s tabuľkami, zabalené do špeciálneho pevného, používateľského rozhrania. V praxi sa objem kódu "150 človekorokov" mení skôr na záťaž ako na konkurenčnú výhodu. Čím viac kódu - tým vyššie náklady na podporu, tým ľahšie sa prispôsobuje novým podmienkam a tým vyšší je vstupný prah pre nových vývojárov a klientov.

Mnohé modulárne stavebné systémy dnes pripomínajú ľahkopádne a zastarané "frankensteinovské konštrukcie", kde každá neopatrná zmena môže viesť k poruchám. Každý nový modul zvyšuje zložitosť už aj tak preťaženého systému a mení ho na labyrinth, ktorému rozumie len niekoľko špecialistov, čo ešte viac sťaže jeho údržbu a modernizáciu.

Zložitosť si uvedomujú aj samotní vývojári, ktorí sa pravidelne pozastavujú pri refaktorovaní - revízii architektúry, aby zohľadnili vznik nových technológií. Avšak aj keď sa refactoring vykonáva pravidelne, zložitosť nevyhnutne rastie. Architekti takýchto systémov si na rastúcu zložitosť zvyknú, ale pre nových používateľov a špecialistov sa stáva neprekonateľnou prekážkou. Výsledkom je, že všetky odborné znalosti sa sústredia v rukách niekoľkých vývojárov a systém prestáva byť škálovateľný. Z krátkodobého hľadiska sú takito odborníci užitoční, ale z dlhodobého hľadiska sa stávajú súčasťou problému.

*Organizácie budú naďalej integrovať "malé" údaje s ich veľkými dátami a je hlúpe veriť, že jedna aplikácia - akokoľvek drahá alebo robustná - zvládne všetko [109].*

- Phil Simon, moderátor podcastu Conversations About Collaboration

Vzniká oprávnená otázka: naozaj potrebujeme také ťažkopádne a uzavreté systémy na výpočet nákladov a časového rozvrhu práce vo forme tabuľiek, keď iné odvetvia už dávno zvládajú podobné úlohy pomocou analytických nástrojov s otvorenými údajmi a transparentnou logikou?

V súčasnosti sú v stavebnictve stále žiadane uzavreté modulárne platformy, a to predovšetkým kvôli špecifickám účtovania nákladov (obr. 5.1-7). Takéto systémy sa často používajú na prevádzkovanie sivých alebo neprehľadných systémov, ktoré umožňujú, aby sa skutočné náklady pred zákazníkom skryli. Keďže však odvetvie digitálne dozrieva, predovšetkým zákazníci, a prechádza do takzvanej "uberizovanej éry", sprostredkovatelia, konkrétnie stavebné spoločnosti so svojimi ERP, stratia svoj význam pri výpočte času a nákladov. To navždy zmení tvár stavebníctva. Viac sa dočítate v poslednej časti knihy a v kapitole "Stavebníctvo 5.0: Ako zarobiť peniaze, keď sa už nedá skryť".

Tisíce starších riešení nahromadených za posledných 30 rokov s tisícami človekorokov investovaných do vývoja začnú rýchlo miznúť. Prechod na otvorenú, transparentnú a flexibilnú správu údajov je nevyhnutný. Jedinou otázkou je, ktoré spoločnosti sa týmto zmenám dokážu prispôsobiť a ktoré zostanú rukojemníkmi starého modelu.

Podobná situácia je aj v oblasti nástrojov CAD (BIM), ktorých údaje dnes napĺňajú objemové parametre projektových jednotiek v systémoch ERP/PMIS. Pôvodne myšlienka BIM (vyvinutá ešte v roku 2002 [110]) vychádzala z koncepcie jednej integrovanej databázy, ale v praxi si dnes práca s BIM vyžaduje celý súbor špecializovaného softvéru a formátov. To, čo malo zjednodušiť projektovanie a riadenie výstavby, sa zmenilo na ďalšiu vrstvu proprietárnych riešení, ktoré komplikujú integráciu a znížujú flexibilitu podnikania.

## Ďalšie kroky: efektívne využívanie údajov o projektoch

V tejto časti sme ukázali, ako sa štruktúrované údaje stávajú základom pre presné výpočty nákladov a harmonogramu stavebných projektov. Automatizácia procesov QTO, plánovania a odhadu znižuje náklady na prácu a výrazne zvyšuje presnosť výsledkov.

Na zhrnutie tejto časti je vhodné zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované prístupy na vaše každodenné úlohy. Tieto prístupy sú univerzálne - sú užitočné tak pre digitálnu transformáciu spoločnosti, ako aj pre každodennú prácu odborníkov zapojených do výpočtov:

■ Automatizácia rutinných výpočtov

- Pokúste sa nájsť štandardné kalkulácie nákladov na prácu, s ktorými sa môžete pri svojej práci stotožniť.
- Analyzujte, ktoré metódy sa používajú na výpočet nákladov alebo prác alebo procesov na stavenisku vo vašej krajine (obrázok 5.1-7).
- Ak pracujete so systémom CAD - naučte sa funkcie automatického získavania špecifikácií a údajov QTO vo vašom softvéri CAD (BIM-).
- Používanie LLM na písanie návrhu kódu na automatizáciu výpočtov

■ Vývoj vlastných nástrojov pre QTO

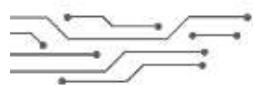
- Vytvorenie skriptov alebo tabuľiek na automatizáciu počítania objemu
- Štandardizácia kategórií a skupín prvkov pre konzistentný prístup k hodnoteniu
- zdokumentovať metodiku výpočtu s cieľom zabezpečiť reprodukovateľnosť výsledkov v nových projektoch

■ Zapojte do svojej práce rôzne aspekty projektu

- Ak pracujete s modulárnymi systémami, skúste vizualizovať svoje procesy nielen ako diagramy alebo grafy, ale aj na úrovni údajov - najmä vo forme tabuľiek.
- Ovládanie automatického spájania údajov získaných z databáz CAD s výpočtami - pomocou kódu Python s využitím zoskupovania, filtrovania a agregácie
- Vytváranie prehľadných vizualizácií skupín QTO na prezentáciu komplexných informácií kolegom a zákazníkom

Tieto kroky pomôžu vybudovať udržateľný systém výpočtu založený na automatizácii a štandardizácii údajov. Tento prístup zlepší presnosť a zníži rutinu každodenných problémov s výpočtom.

Nasledujúce kapitoly sa zameriavajú na technické aspekty produktov CAD - (BIM-) a na dôvody, prečo sa databázy CAD stále ľahko integrujú do podnikových procesov. Ak vás teraz nezaujíma história implementácie BIM v stavebnictve, vývoj CAD nástrojov a technické aspekty práce s týmito technológiami, môžete prejsť rovno na siedmu časť knihy "Rozhodovanie založené na údajoch".



## MAXIMÁLNE POHODLIE S TLAČENOU VERZIOU

Držíte v rukách bezplatnú digitálnu verziu knihy **Data-Driven Construction**. Pre pohodlnejšiu prácu a rýchly prístup k materiálom vám odporúčame venovať pozornosť **tlačenému vydaniu**:



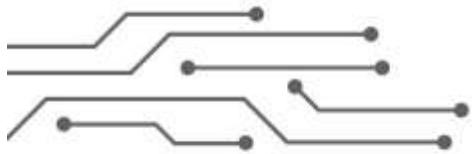
■ **Vždy po ruke:** kniha v tlačenej podobe sa stane spoľahlivým pracovným nástrojom, ktorý vám umožní rýchlo nájsť a použiť správne vizualizácie a diagramy v akejkoľvek pracovnej situácii

■ **Vysoká kvalita ilustrácií:** všetky obrázky a grafika v tlačenom vydani sú prezentované v maximálnej kvalite

■ **Rýchly prístup k informáciám:** jednoduchá navigácia, možnosť robiť si poznámky, záložky a pracovať s knihou kdekoľvek.

Zakúpením plnej tlačenej verzie knihy získate praktický nástroj na pohodlnú a efektívnu prácu s informáciami: možnosť rýchlo využívať vizuálne materiály pri každodenných úlohách, rýchlo nájsť potrebné schémy a robiť si poznámky. Okrem toho svojou kúpou podporíte šírenie otvorených poznatkov.

Tlačenú verziu knihy si môžete objednať na adrese: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## VI ČASŤ

# CAD A BIM: MARKETING, REALITA A BUDÚCNOSŤ PROJEKTOVÝCH ÚDAJOV V STAVEBNÍCTVE

Šiesta časť knihy predstavuje kritickú analýzu vývoja CAD a BIM -technológií a ich vplyvu na procesy správy údajov v stavebníctve. Sleduje historickú premenu koncepcie BIM od pôvodnej myšlienky integrovanej databázy až po súčasné marketingové konštrukcie presadzované výrobcami softvéru. Hodnotí sa vplyv proprietárnych formátov a uzavretých systémov na efektívnosť spracovania projektových údajov a celkovú výkonnosť stavebníctva. Podrobne sa analyzujú problémy kompatibility rôznych CAD-systémov a ťažkosti ich integrácie s obchodnými procesmi stavebných spoločností. Diskutuje sa o súčasných trendoch smerujúcich k zjednodušeným otvoreným dátovým formátom, ako je napríklad USD, a o ich potenciálnom vplyve na odvetvie. Uvádzajú sa alternatívne prístupy k získavaniu informácií z uzavretých systémov vrátane techník reverzného inžinierstva. Analyzujú sa vyhliadky na uplatnenie umelej inteligencie a strojového učenia na automatizáciu procesov projektovania a analýzy údajov v stavebníctve. Formulujú sa prognózy vývoja projekčných technológií orientovaných na skutočné potreby používateľov, a nie na záujmy dodávateľov softvéru.

## KAPITOLA 6.1.

# VZNIK KONCEPCIÍ BIM- V STAVEBNÍCTVE

Pôvodne táto šiesta časť, venovaná CAD (BIM), nebola v prvej verzii knihy zahrnutá. Témy proprietárnych formátov, geometrických jadier a uzavretých systémov sú príliš odborné, preplnené detailmi a zdanivo nepoužiteľné pre tých, ktorí chcú len pochopiť, ako pracovať s údajmi. Spätná väzba a žiadosti o doplnenie objasnení do prvej verzie knihy však ukázali, že bez pochopenia zložitosti vnútorného fungovania systémov CAD, geometrických jadier, rôznorodosti formátov a nekompatibilných schém ukladania tých istých údajov nie je možné skutočne pochopiť, prečo koncepcie presadzované výrobcami často stážujú prácu s informáciami a bránia prechodu na otvorené parametrické navrhovanie. Preto táto časť zaujala v štruktúre knihy samostatné miesto. Ak pre vás CAD (BIM) nie je prioritou, môžete preskočiť rovno na ďalšiu časť - "ČASŤ VII: Rozhodovanie založené na údajoch, analytika, automatizácia a strojové učenie".

### **História vzniku BIM a otvoreného BIM ako marketingových konceptov dodávateľov CAD-**

S príchodom digitálnych údajov v 90. rokoch 20. storočia sa počítačové technológie zaviedli nielen do obchodných procesov, ale aj do procesov projektovania, čo viedlo k vzniku pojmov ako CAD (systémy počítačom podporovaného projektovania) a neskôr BIM (informačné modelovanie budov)

Ako každá inovácia však nie sú konečným bodom vývoja. Koncepty ako BIM sa stali dôležitým milníkom v histórii stavebníctva, ale skôr či neskôr môžu ustúpiť lepším nástrojom a prístupom, ktoré budú lepšie spĺňať výzvy budúcnosti.

Koncepcia BIM, ktorá sa objavila v roku 2002 a ktorá je preťažená vplyvom dodávateľov CAD a zmätená zložitosťou vlastnej implementácie, sa možno nedožije svojho tridsiateho výročia, podobne ako rocková hviezda, ktorá jasne zažiarila, ale rýchlo zhasla. Dôvod je jednoduchý: požiadavky dátových vedcov sa menia rýchlejšie, než sa im dokážu prispôsobiť dodávatelia CAD.

Dnešní odborníci v stavebníctve, ktorí čelia nedostatku kvalitných údajov, požadujú interoperabilitu medzi platformami a prístup k otvoreným údajom z projektov CAD-, aby sa zjednodušila ich analýza a spracovanie. Zložitosť údajov CAD a zmätočné spracovanie údajov CAD má negatívny vplyv na všetkých účastníkov stavebného procesu: projektantov, projektových manažérov, stavebných pracovníkov na stavbe a v konečnom dôsledku aj na klienta.

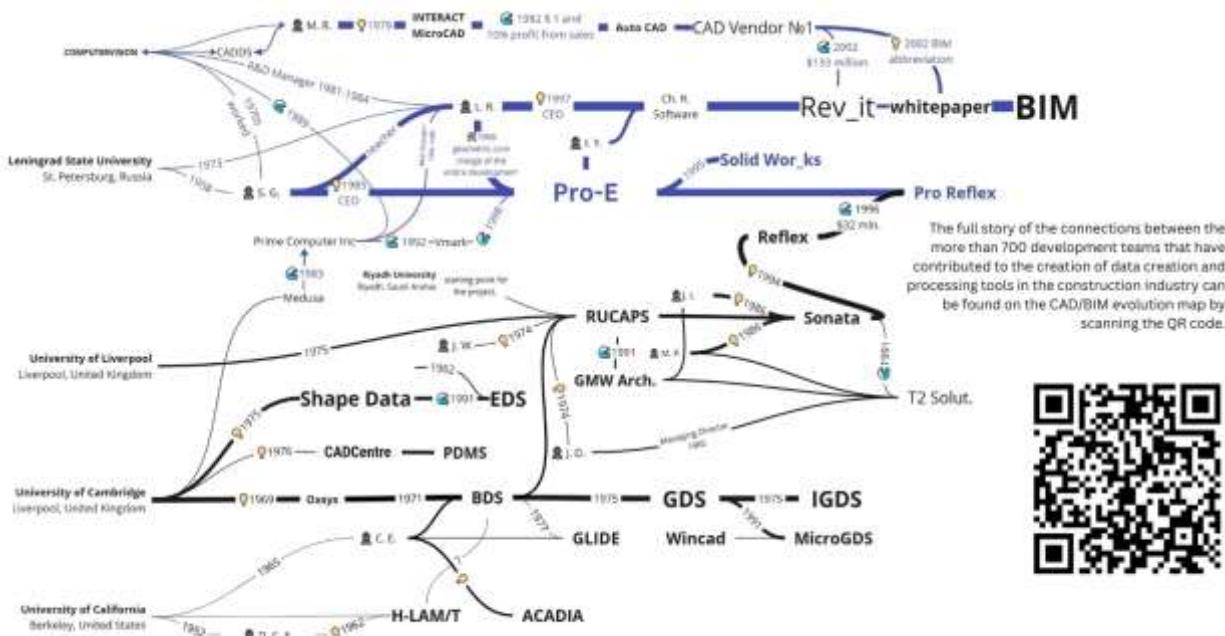
Namiesto plnohodnotného súboru údajov na prevádzku dnes zákazník a investor dostávajú kontajnery vo formátoch CAD-, ktoré si vyžadujú zložité geometrické jadrá, pochopenie dátových schém, každoročne aktualizovanú dokumentáciu API a špecializovaný softvér CAD (BIM) na prácu s údajmi. Zároveň veľká časť údajov o projektovaní zostáva nevyužitá.

V dnešnom svete projektovania a výstavby vedie zložitosť prístupu k údajom CAD k nadmernému

inžinierstvu pri riadení projektov. Stredné a veľké spoločnosti pracujúce s údajmi CAD alebo vyvíjajúce BIM -riešenia sú bud' nútené udržiavať úzke vzťahy s dodávateľmi CAD riešení na prístup k údajom prostredníctvom API, alebo obchádzať obmedzenia dodávateľov CAD pomocou drahých konvertorov SDK na reverzné inžinierstvo, aby získali otvorené údaje [75].

Vlastný prístup k údajom je zastaraný a už nespĺňa požiadavky dnešného digitálneho prostredia. Budúcnosť rozdelí spoločnosti na dva typy: tie, ktoré efektívne využívajú otvorené údaje, a tie, ktoré trh opustia.

Pojem BIM (Building Information Modeling), sa v stavebníctve objavil s publikáciou jedného z hlavných dodávateľov CAD - Whitepaper BIM [54] v roku 2002 a doplnený strojárskym pojmom BOM (Bills of Materials), vznikol z parametrického prístupu k tvorbe a spracovaniu projektových údajov (obr. 6.1-1). Parametrický prístup k tvorbe a spracovaniu projektových údajov bol ako jeden z prvých implementovaný v systéme Pro-E pre strojárske projektovanie (MCAD). Tento systém sa stal prototypom [111mnohých moderných riešení CAD vrátane tých, ktoré sa dnes používajú v stavebníctve.



*Informačné modelovanie budov je stratégia..... [názov spoločnosti dodávajúcej CAD] na aplikáciu informačných technológií v stavebníctve.*

- Biela kniha BIM, 2002 [60].

V polovici roku 2000 začali "výskumníci" prepájať koncepciu BIM-, ktorú v roku 2002 zverejnili dodávateľ CAD-, s predchádzajúcimi vedeckými prácam, ako napríklad BDS Charlesa Eastmana , ktoré sa stali základom pre systémy ako GLIDE, GBM, BPM, RUCAPS. Charles Eastman vo svojej priekopníckej práci Building Description System (1974) položil teoretické základy moderného informačného modelovania. Termín "databáza " sa v jeho práci objavuje 43-krát (obr. 6.1-2) - častejšie ako ktorýkoľvek iný, s výnimkou slova "budova".

Eastmanova kľúčová myšlienka spočívala v tom, že všetky informácie o budove - od geometrie až po vlastnosti prvkov a ich vzájomné vzťahy - by mali byť uložené v jednej štruktúrovanej databáze. Práve z tejto databázy možno automaticky generovať a analyzovať výkresy, špecifikácie, výpočty a súlad s predpismi. Eastman výslovne kritizoval výkresy ako zastaraný a nadbytočný spôsob komunikácie, pričom poukazoval na duplicitu informácií, problémy s aktualizáciou a potrebu ručnej aktualizácie pri zmenách. Namiesto toho navrhhol jediný digitálny model v databáze, kde sa každá zmena vykoná raz a automaticky sa premietne do všetkých zobrazení.

Je pozoruhodné, že Eastman vo svojej koncepcii nekladie vizualizáciu do centra pozornosti. V centre jeho systému boli informácie: parametre, vzťahy, atribúty, možnosti analýzy a automatizácie. Výkresy v jeho chápání boli len jednou z foriem zobrazenia údajov z databázy, nie primárnym zdrojom informácií o návrhu.

V prvej Bielej knihe BIM od popredného dodávateľa CAD sa výraz "databáza " používal rovnako často ako v BDS Charlesa Eastmana - 23-krát [60] siedmich stranach a bol jedným z najobľúbenejších slov v dokumente po slovách "budova", "informácie", "modelovanie" a "návrh". Do roku 2003 sa však výraz "databáza" v podobných dokumentoch objavil len dvakrát [61] a koncom roka 2000 sa téma databáz z diskusie o údajoch o projektovaní prakticky vytratila. Výsledkom bolo, že koncepcia "jednotnej integrovanej databázy pre vizuálnu a kvantitatívnu analýzu" nebola nikdy úplne realizovaná.

Stavebníctvo tak prešlo od progresívnej koncepcie BDS Charlesa Eastmana s dôrazom na databázy a myšlienky Samuela Geisberga o automatickej aktualizácii konštrukčných údajov z databáz v strojárskom produkte Pro-E (predchadca populárnych CAD -riešení používaných v stavebníctve v súčasnosti) až po súčasný marketingový BIM, kde sa správa údajov prostredníctvom databáz takmer nespomína, napriek tomu, že práve táto koncepcia bola pôvodným teoretickým



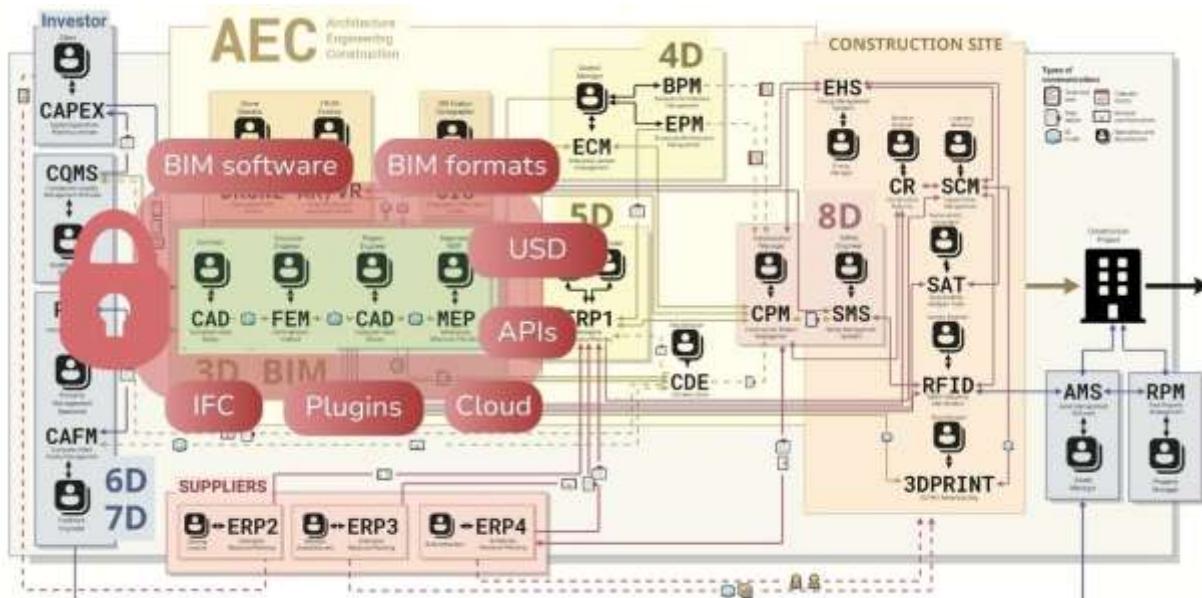
výmeny informácií sú používateľia nútení platiť za predplatné a používať proprietárne API.

- **Údaje patria predajcom, nie používateľom.** Informácie o projektoch sú uzamknuté v proprietárnych formátoch alebo clouдовých službách, namiesto toho, aby boli k dispozícii v otvorených a nezávislých formátoch.

Konštruktéri a projektoví manažéri často nemajú prístup k databáze CAD -systémom, ani k formátu, v ktorom sú uložené ich vlastné projektové údaje. To znemožňuje rýchle overenie informácií alebo formulovanie požiadaviek na štruktúru a kvalitu údajov (obr. 6.1-3). Prístup k takýmto údajom si vyžaduje celý súbor špecializovaných programov prepojených prostredníctvom API a zásuvných modulov, čo viedie k nadmernej byrokratizácii procesov v stavebníctve. Pritom tieto údaje súčasne využívajú desiatky informačných systémov a stovky špecialistov.

*Musíme byť schopní spravovať všetky tieto údaje [CAD (BIM)], ukladať ich digitálne a predávať softvér na riadenie životného cyklu a procesov, pretože **na každého inžiniera [projektanta]**, ktorý niečo vytvorí [v softvéri CAD], **pripadá desať ľudí, ktorí s týmito údajmi pracujú**" [41].*

- Generálny riaditeľ spoločnosti CAD - dodávateľa, ktorý vytvoril koncept BIM, 2005.



Obr. 6.1-3 CAD- (BIM-) databázy zostávajú jedným z posledných uzavretých systémov pre IT oddelenia a správcov údajov v ekosystéme stavebných podnikov.

Ked' je zrejmé, že BIM je skôr prostriedkom na komericalizáciu databáz ako plnohodnotným nástrojom na správu databáz, vyvstáva logická otázka: ako získať kontrolu nad údajmi? Odpoveďou je použitie otvorených dátových štruktúr, v ktorých sa vlastníkom informácií stáva používateľ, a nie dodávateľ softvéru.

Používatelia a vývojári riešení v stavebníctve, podobne ako ich kolegovia v iných odvetviach, sa nevyhnutne odklonia od nejasnej terminológie dodávateľov softvéru, ktorá dominovala posledných 30 rokov, a zamerajú sa na kľúčové aspekty digitalizácie - "údaje" a "procesy".

Ešte koncom 80. rokov 20. storočia sa kľúčová oblasť digitálneho rozvoja v stavebníctve prezentovala ako otázka prístupu k údajom a riadenia projektových informácií. Postupom času sa však ľažisko zmenilo. Namiesto rozvoja transparentných a prístupných prístupov k práci s údajmi sa aktívne presadzoval formát IFC a otvorená koncepcia BIM ako pokusy odviesť pozornosť odborníkov od tém správy projektových databáz.

## Vznik otvoreného formátu IFC v stavebníctve

Takzvaný otvorený formát IFC (Industry Foundation Classes) je umiestnený ako štandard na zabezpečenie interoperability medzi rôznymi systémami CAD (BIM -). Jeho vývoj prebiehal v rámci organizácií, ktoré boli vytvorené a riadené hlavnými dodávateľmi CAD. Na základe formátu IFC dve spoločnosti CAD- v roku 2012 vytvorili marketingový koncept OPEN BIM[63].

**IFC** (Industry Foundation Classes) je otvorený štandard na výmenu údajov v stavebníctve, ktorý je určený na zabezpečenie interoperability medzi rôznymi systémami CAD - (BIM-).

Koncept Open BIM - zahŕňa prácu s informáciami z databáz CAD a výmenu informácií medzi systémami prostredníctvom otvoreného formátu na výmenu údajov CAD - IFC.

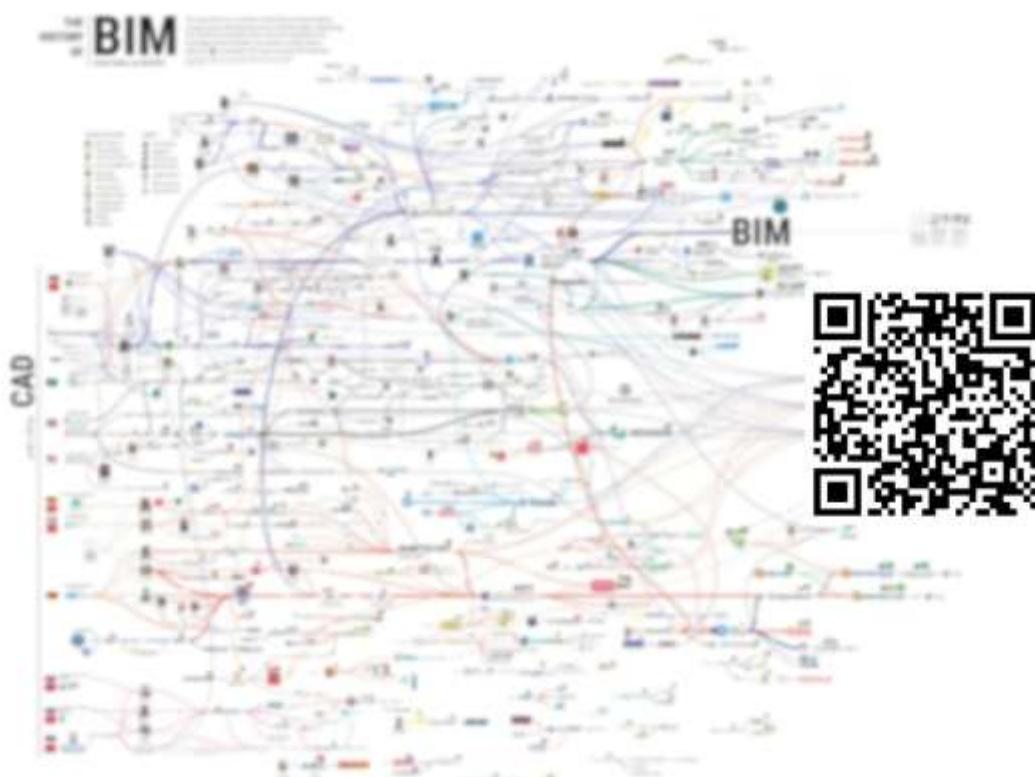
*Program Open BIM je marketingová kampaň iniciovaná... [1 dodávateľ CAD]... [2 dodávateľ CAD] a ďalšími spoločnosťami s cieľom podporiť a uľahčiť globálnu koordinovanú propagáciu koncepcie OPEN BIM v celom odvetví AEC, pričom účastníci programu majú k dispozícii jednotnú komunikáciu a spoločnú značku.*

- Z webovej stránky dodávateľa CAD, program OPEN BIM, 2012 [113].

IFC bol koncom 80. rokov 20. storočia adaptovaný Technickou univerzitou v Mnichove zo strojárskeho formátu STEP, neskôr ho zaregistrovala veľká konštrukčná spoločnosť a významný dodávateľ CAD- a v roku 1994 vytvorili IAI (Industry Alliance for Interoperability) [114] (obr. 6.1-4). Formát IFC bol vyvinutý na zabezpečenie interoperability medzi rôznymi systémami CAD a vychádzal zo zásad stanovených v strojárskom formáte STEP, ktorý zasa vznikol z formátu IGES vytvoreného v roku 1979 skupinou používateľov a dodávateľov CAD s podporou NIST (The National Institute of Standards and

Technology) a Ministerstva obrany USA [115].

Zložitá štruktúra IFC, jeho úzka závislosť od geometrického jadra, ako aj rozdiely v implementácii formátu rôznymi softvérovými riešeniami však viedli k mnohým problémom pri jeho praktickom používaní. S podobnými ťažkoťami - stratou detailov, obmedzením presnosti a potrebovou používať prechodné formáty - sa predtým stretávali odborníci na strojárstvo pri práci s formátmi IGES a STEP, z ktorých vznikol IFC.



Obr. 6.1-4 Mapa prepojení medzi vývojovými tímami a produktmi CAD (BIM) [116].

V roku 2000 ten istý dodávateľ CAD, ktorý zaregistroval formát IFC a vytvoril organizáciu IAI (neskôr bS), vydáva bielu knihu "Integrated Design and Manufacturing: Benefits and Rationale". [65]. V dokumente sa zdôrazňuje dôležitosť zachovania úplnej granularity údajov pri výmene medzi programami v rámci tohto systému bez použitia neutrálnych formátov, ako je IGES, STEP [identický s IFC]. Namiesto toho sa navrhovalo, aby aplikácie mali priamy prístup k základnej databáze CAD, aby sa zabránilo strate presnosti informácií.

V roku 2002 ten istý dodávateľ CAD kupuje parametrický produkt BOM (obr. 3.1-18, podrobnejšie v tretej časti) a na jeho základe vytvára koncepciu BIM. V dôsledku toho sa pri výmene údajov o stavebných projektoch používajú len uzavreté formáty CAD alebo formát IFC (STEP), o ktorého obmedzeniach písal v roku 2000 sám dodávateľ CAD, ktorý tento formát do stavebníctva priniesol.

Podrobná história interakcie viac ako 700 vývojových tímov, ktoré sa podielali na tvorbe nástrojov na

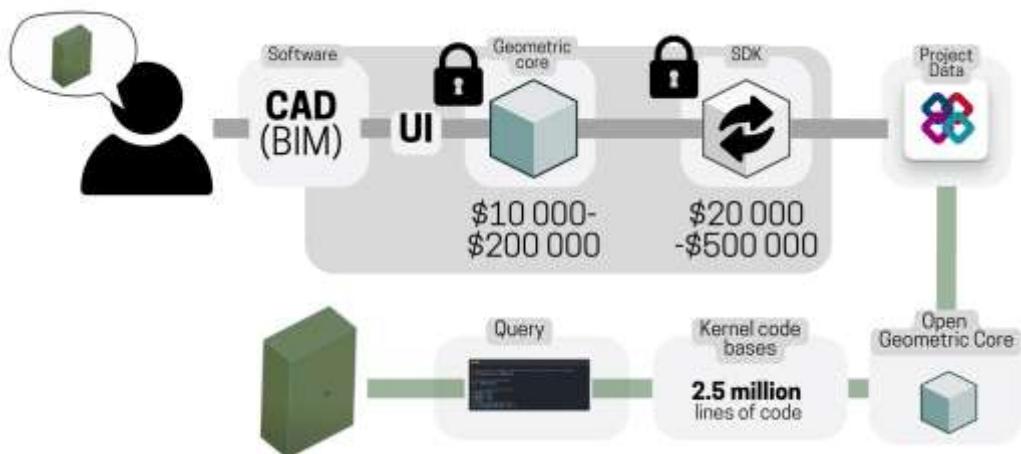
vytváranie a spracovanie stavebných údajov, je uvedená v mape "Vývoj CAD (BIM)" [116]. [116].

Otvorený formulár IFC pozostáva z geometrického opisu konštrukčných prvkov a opisu metainformácií. Na reprezentáciu geometrie vo formáte IFC sa používajú rôzne metódy, napríklad CSG a Swept Solids: vedúcim štandardom na prenos geometrie prvkov vo formáte IFC sa však stala parametrická reprezentácia BREP, pretože tento formát je podporovaný pri exporte z programov CAD-(BIM-) a umožňuje prípadné úpravy prvkov pri spätnom importe IFC do programov CAD.

## Problém formátu IFC v závislosti od geometrického jadra

Vo väčšine prípadov, keď je geometria v IFC definovaná parametricky (BREP), je nemožné vizualizovať alebo načítať geometrické vlastnosti, ako je objem alebo plocha projektových entít, len pomocou súboru IFC, pretože na prácu s geometriou a jej vizualizáciu je v tomto prípade potrebné jadro geometrie (obrázok 6.1-5), ktoré na začiatku chýba.

**Geometrické jadro** je softvérový komponent, ktorý poskytuje základné algoritmy na vytváranie, úpravu a analýzu geometrických objektov v CAD (CAD), BIM a iných inžinierskych aplikáciach. Je zodpovedné za konštrukciu 2D a 3D -geometrie, ako aj za operácie s ňou, ako sú: booleovské operácie, vyhladzovanie, priesčníky, transformácie a vizualizácia.



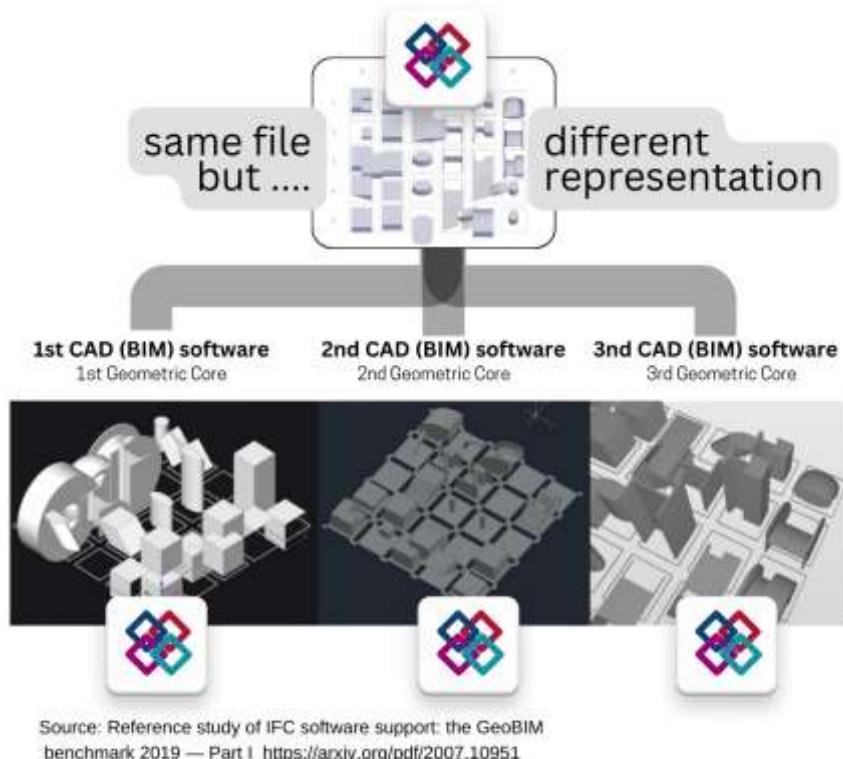
Obrázok 6.1-5 Vytváranie geometrie prostredníctvom CAD- softvéru dnes prebieha prostredníctvom proprietárnych geometrických jadier a SDK, ktoré často nie sú vlastníctvom dodávateľov CAD.

Každý program CAD a všetky programy pracujúce s parametrickými formátmi alebo formátmi IFC majú vlastné alebo zakúpené geometrické jadro. A ak s primitívnymi prvkami vo formáte IFC -BREP nemôžu byť problémy a v programoch s rôznymi geometrickými jadrami sa tieto prvky dajú zobraziť podobne, ale okrem problémov s rôznymi motormi geometrických jadier, existuje dosť prvkov, ktoré majú svoje špecifiká pre správne zobrazenie. Tento problém je podrobne rozpracovaný v medzinárodnej štúdii "A reference study of IFC software support" (Referenčná štúdia softvérovej podpory IFC) uverejnenej v

roku 2019 [117].

Rovnaké štandardizované súbory údajov prinášajú protichodné výsledky, sa našlo len málo spoločných vzorov a zistili sa vážne problémy s podporou štandardu [IFC], pravdepodobne v dôsledku veľmi vysokej zložitosti štandardného modelu údajov. Čiastočne sú tu na vire samotné normy, pretože často ponechávajú niektoré detaily nedefinované, s vysokým stupňom voľnosti a rôznymi možnými interpretáciami. Umožňujú vysokú zložitosť organizácie a ukladania objektov, čo neprispieva k efektívному univerzálnemu porozumeniu, jedinečným implementáciám a konzistentnému modelovaniu údajov [117].

- Referenčná štúdia softvérovej podpory IFC, 2021



Obr. 6.1-6 Rôzne geometrické jadrá poskytujú rôzne reprezentácie tej istej geometrie opísanej parametricky (podľa [117]).

Správne chápanie "niektorých ustanovení" je dostupné plateným členom špeciálnych organizácií, ktoré vypracúvajú IFC. V dôsledku toho sa ten, kto chce získať prístup k dôležitým poznatkom o určitých vlastnostiach IFC, pokúsi spolupracovať s veľkými dodávateľmi CAD- alebo dosiahnuť kvalitatívne posúdenie vlastností vlastným výskumom.

Narazíte na otázku týkajúcu sa importu a exportu údajov prostredníctvom formátu IFC a spýtate sa kolegov predajcov: "Prečo je tak v súbore IFC informácia o parametrickom prenose priestorov? V otvorenej špecifikácii sa o tom nič nehovorí". Odpoveď od "kompetentnejších" európskych dodávateľov: "Áno, nie je to povedané, ale je to povolené".

- Z rozhovoru s vývojárom CAD 2021 [118].

IFC opisuje geometriu prostredníctvom parametrických primitív, ale neobsahuje vložené jadro - jeho úlohu plní program CAD, ktorý zostavuje geometriu prostredníctvom jadra geometrie. Jadro geometrie vykonáva matematické výpočty a definuje priesecníky, zatiaľ čo IFC poskytuje len údaje na ich interpretáciu. Ak IFC obsahuje nesprávne plochy, rôzne programy s rôznymi geometrickými jadrami ich môžu v závislosti od jadra buď ignorovať, alebo vytvárať chyby.

Výsledkom je, že pre prácu s formátom IFC je potrebné zodpovedať hlavnú otázku, na ktorú je ťažké nájsť jednoznačnú odpoveď - aký nástroj, s akým geometrickým jadrom použiť, aby sa dosiahla kvalita údajov, ktorú mal projekt pôvodne v CAD programe, z ktorého bol IFC získaný?

Problémy s kvalitou údajov a zložitosť formátu IFC neumožňujú priame použitie údajov o návrhu na automatizáciu procesov, analýzu a spracovanie údajov, čo často vedie vývojárov k nevyhnutej potrebe používať uzavreté CAD -riešenia s "kvalitným" prístupom k údajom [63], o čom písal sám dodávateľ, ktorý IFC zaregistroval v roku 1994 [65].

Všetky osobitosti mapovania a generovania parametrov IFC v geometrickom jadre môžu realizovať len veľké tímy vývojárov, ktorí majú skúsenosti s prácou s geometrickými jadrami. Preto je súčasná prax osobitostí a zložitosťi formátu IFC výhodná predovšetkým pre dodávateľov CAD- a má veľa spoločného so stratégou veľkých dodávateľov softvéru "adoptovať, rozšíriť, zničiť", keď rastúca zložitosť štandardu v skutočnosti vytvára prekážky pre malých hráčov na trhu [94].

Stratégia veľkých predajcov pri takejto stratégii môže byť prispôsobenie otvorených štandardov, pridanie proprietárnych rozšírení a funkcií s cieľom vytvoriť závislosť používateľov na svojich produktoch a následne vytlačiť konkurenciu.

Formát IFC, ktorý má byť univerzálnym mostom medzi rôznymi systémami CAD- (BIM-), v skutočnosti plní úlohu indikátora problémov s kompatibilitou medzi geometrickými jadrami rôznych platform CAD, podobne ako formát STEP, z ktorého pôvodne vznikol.

V dôsledku toho je dnes úplná a kvalitná implementácia ontológie IFC realizovateľná pre veľkých dodávateľov CAD, ktorí môžu investovať značné zdroje na podporu všetkých entít a ich mapovanie na vlastné interné jadro geometrie, ktoré pre IFC ako štandard neexistuje. Veľkí dodávateľia majú tiež možnosť koordinovať medzi sebou technické detaile prvkov, ktoré nemusia byť dostupné ani najaktívnejšiemu účastníkovi organizácií využívajúcich formát IFC.

Pre malé nezávislé tímy a projekty s otvoreným zdrojovým kódom, ktoré sa snažia podporovať vývoj interoperabilných formátov, sa absencia vlastného geometrického jadra stáva vážnym problémom. Bez neho je prakticky nemožné zohľadniť všetky mnohé jemnosti a nuansy spojené s výmenou údajov medzi platformami.

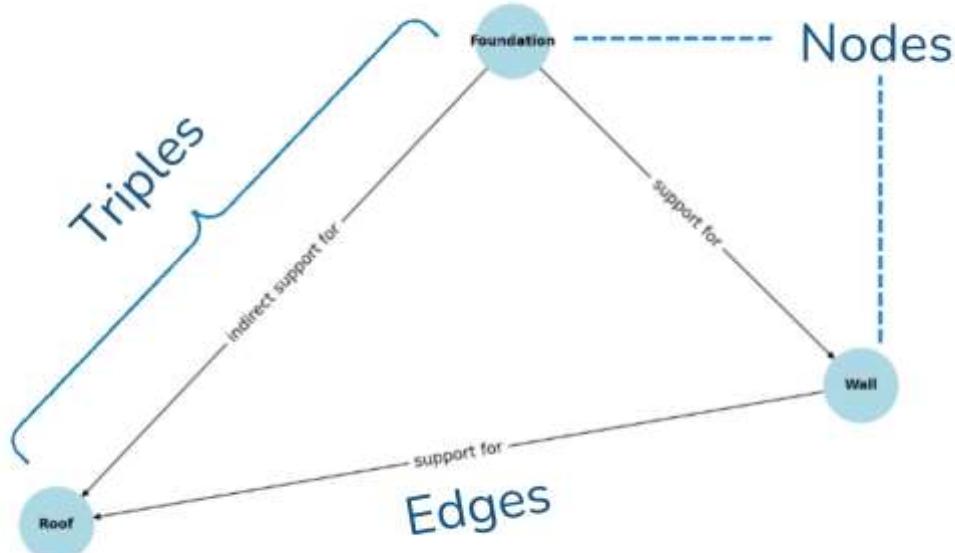
S rozvojom parametrického formátu IFC a otvorenej koncepcie BIM sa v stavebníctve zintenzívnil diskusie o úlohe ontológie a sémantiky pri správe údajov a procesov.

## Výskyt v konštrukcii témy sémantiky a ontológie

Vďaka myšlienkom sémantického internetu konca 90. rokov 20. storočia a úsiliu organizácií zapojených do vývoja formátu IFC, sa sémantika a ontológie stali jedným z kľúčových prvkov normalizácie, o ktorej sa v stavebníctve diskutuje do polovice 20. storočia.

Sémantické technológie predstavujú zjednotenie, štandardizáciu a úpravu veľkých polí heterogénnych údajov a realizáciu komplexného vyhľadávania.

Na ukladanie sémantických údajov sa používa jazyk OWL (Web Ontology Language), reprezentovaný ako grafy RDF -triplets (Resource Description Framework) (obr. 6.1-7). OWL sa vzťahuje na grafové dátové modely, ktorých typy sme podrobnejšie rozobrali v kapitole "Dátové modely: dátové vzťahy a vzťahy medzi prvkami".



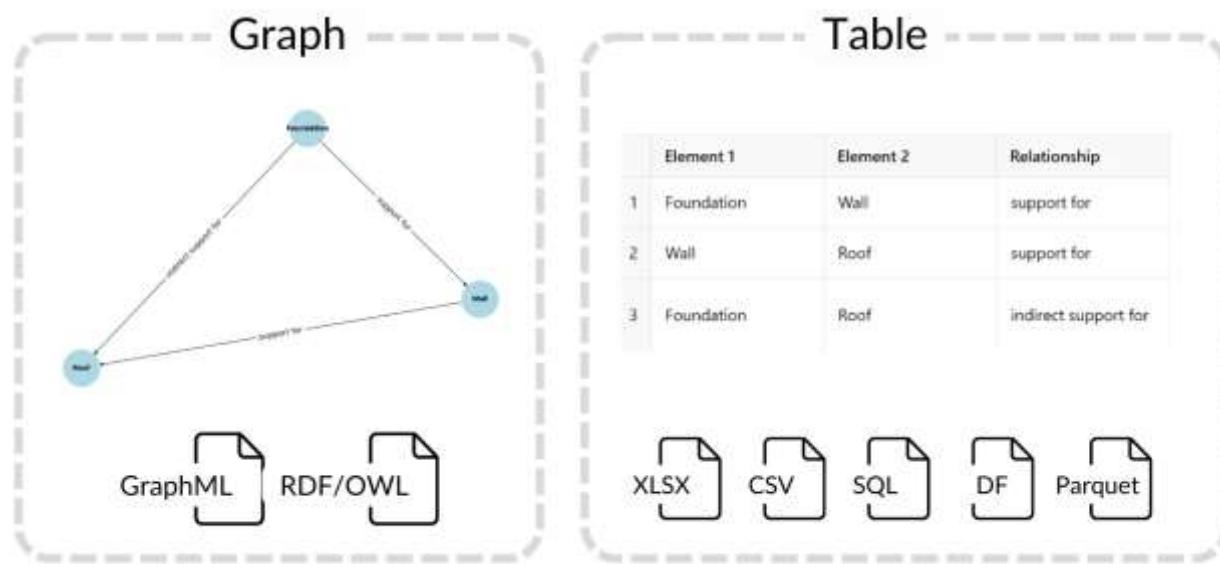
Obr. 6.1-7 Dátový model RDF: Uzly, hrany a trojice znázorňujúce vzťahy medzi stavebnými prvkami.

Logické odvodzovanie risonerov (softvér na automatické logické odvodzovanie) teoreticky umožňuje odvodzovať nové výroky z ontológií. Napríklad, ak ontológia budovy zaznamenáva, že "základ je podperou pre stenu" a "stena je podperou pre strechu" (obr. 6.1-7), risoner dokáže automaticky odvodiť, že "základ je podperou pre strechu".

Takýto mechanizmus je užitočný na optimalizáciu analýzy údajov, pretože sa vyhýba explicitnému predpisovaniu všetkých závislostí. Nevytvára však nové poznatky, ale len identifikuje a štruktúruje už známe fakty.

Sémantika sama o sebe nevytvára nový význam alebo znalosti a v tomto ohľade nie je nadradená iným technológiám ukladania a spracovania údajov. Reprezentácia údajov z relačných databáz ako trojíc ich nerobí zmysluplnnejšími. Nahradenie tabuľiek grafovými štruktúrami môže byť užitočné na zjednotenie dátových modelov, jednoduché vyhľadávanie a bezpečné úpravy, ale nerobí údaje "inteligentnejšími" - počítač nezačne lepšie rozumieť ich obsahu.

Logické vzťahy v údajoch možno organizovať bez zložitých sémantických technológií (obr. 6.1-8). Tradičné relačné databázy (SQL), ako aj formáty CSV alebo XLSX umožňujú vytvárať podobné závislosti. Napríklad v stĺpcovej databáze môžete pridať pole "podpora strechy" a pri vytváraní steny automaticky priradiť strechu k základom. Tento prístup je implementovaný bez použitia RDF, OWL, grafov alebo risonerov, pričom zostáva jednoduchým a efektívny riešením na ukladanie a analýzu údajov.



Obr. 6.1-8 Porovnanie dátových modelov grafov a tabuliek na reprezentáciu rovnakých logických vzťahov.

Rozhodnutie viacerých veľkých stavebných spoločností a organizácie pre vývoj formátu IFC [94] nasledovalo koncepciu sémantického webu, ktorá sa koncom 90. rokov 20. storočia zdala sľubná, malo významný vplyv na vývoj noriem v stavebníctve.

Paradoxom však je, že samotná koncepcia sémantického webu, pôvodne určená pre internet, nebola široko prijatá ani vo svojom pôvodnom prostredí. Napriek rozvoju RDF a OWL sa plnohodnotný sémantický web v pôvodnej koncepcii neobjavil a jeho vznik je už málo pravdepodobný.

## Prečo sémantické technológie nespĺňajú očakávania v stavebníctve

Iné odvetvia sa stretli s obmedzeniami technológií na využívanie sémantiky. V hernom priemysle sa pokusy o opis herných objektov a ich interakcií prostredníctvom ontológií ukázali ako neúčinné z dôvodu vysokej dynamiky zmien. V dôsledku toho sa uprednostnili jednoduchšie formáty údajov, ako sú XML a JSON, spolu s algoritmickými riešeniami. Podobná situácia bola v sektore nehnuteľností: vzhľadom na regionálne rozdiely v terminológii a časté zmeny na trhu sa používanie ontológií ukázalo ako príliš zložité, zatiaľ čo jednoduché databázy a štandardy, ako napríklad RETS [119], sa dokázali lepšie vyrovnáť s výzvami výmeny údajov.

Technické ťažkosti, ako napríklad zložitosť značkovania, vysoká práclosť podpory a nízka motivácia vývojárov, spomalili prijatie sémantického webu a v iných odvetviach hospodárstva. RDF (Resource Description Framework) sa nestal masovým štandardom a ontológie sa ukázali ako príliš zložité a ekonomicky neopodstatnené.

V dôsledku toho sa nepodarilo zrealizovať ambicioznu myšlienku vytvorenia globálneho sémantického webu. Hoci niektoré prvky tejto technológie, ako napríklad ontológie a SPARQL, si našli cestu do podnikových riešení, pôvodný cieľ vytvoriť jednotnú komplexnú dátovú štruktúru sa nepodarilo

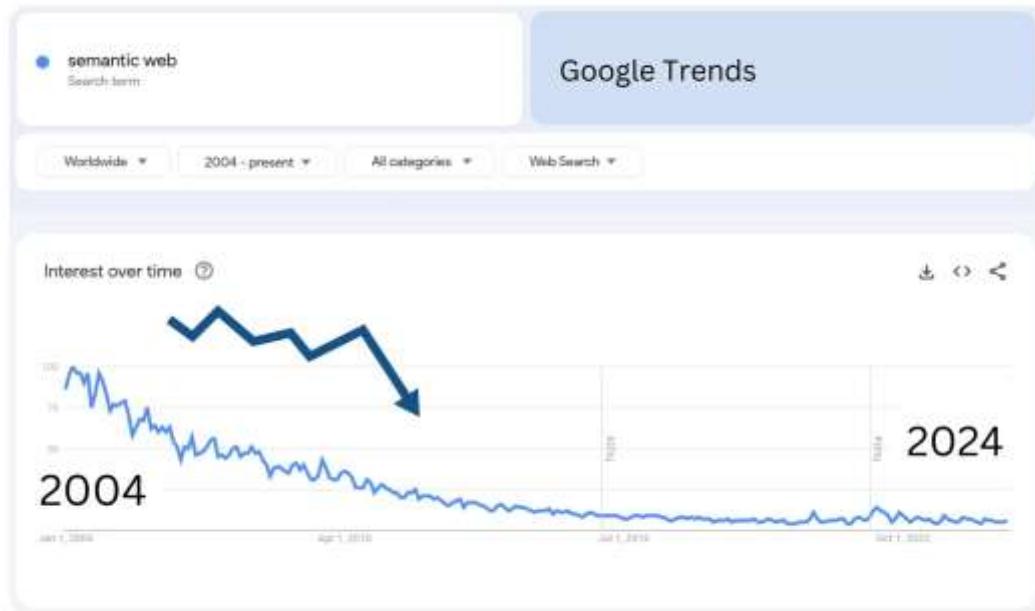
dosiahnuť.

Koncepcia internetu, v ktorom sú počítače schopné dať obsahu zmysel, sa ukázala ako technicky náročná a komerčne nerentabilná. Preto spoločnosti, ktoré podporovali túto myšlienku, nakoniec obmedzili jej používanie na jednotlivé užitočné nástroje a ponechali RDF a OWL pre vysoko špecializované podnikové potreby, a nie pre internet ako celok. Analýza trendov Google (obrázok 6.1-9) za posledných 20 rokov naznačuje, že sémantický web už možno nemá žiadne vyhliadky.

*Nie je potrebné zbytočne množiť subjekty. Ak existuje viacero logicky konzistentných vysvetlení nejakého javu, ktoré ho vysvetľujú rovnako dobre, mali by sme, za rovnakých podmienok, uprednostniť najjednoduchšie z nich.*

britva

Vyvstáva tu logická otázka: prečo vôbec používať pri konštrukcii triplety, risery a SPARQL, keď môžete údaje spracovať pomocou populárnych štruktúrovaných dotazov (SQL, Pandas, Apache®)? V podnikových aplikáciách je SQL štandardom pre prácu s databázami. Naopak, SPARQL si vyžaduje zložité grafové štruktúry a špecializovaný softvér a podľa trendov v Google nepriťahuje záujem vývojárov.



Obr. 6.1-9 Záujem o "sémantické internetové" dotazy podľa štatistik Google.

Grafové databázy a klasifikačné stromy môžu byť v niektorých prípadoch užitočné, ale ich použitie nie je vždy opodstatnené pre väčšinu každodenných úloh. V dôsledku toho má vytváranie znalostných grafov a využívanie technológií sémantického webu zmysel len vtedy, keď je potrebné zjednotiť údaje z rôznych zdrojov alebo realizovať zložité logické závery.

Prechod od tabuľiek ku grafickým dátovým modelom zlepšuje vyhľadávanie a zjednocoje tok informácií, ale nezvyšuje význam údajov pre stroje. Otázkou nie je, či by sa sémantické technológie mali používať, ale v čom majú skutočný význam. Skôr ako vo svojej spoločnosti zavediete ontológiu, sémantiku a grafové databázy, zistite, ktoré spoločnosti už tieto technológie úspešne používajú a kde sa im to nepodarilo.

Napriek ambicioznym očakávaniam sa sémantické technológie nikdy nestali univerzálnym riešením na štruktúrovanie údajov v stavebníctve. V praxi tieto technológie nevedli k univerzálnemu riešeniu, ale len pridali nové zložitosti a tieto snahy sú ozvenou nerealizovaných ambícií koncepcie sémantického internetu, kde očakávania ďaleko prevyšovali realitu.

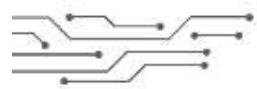


Obr. 6.1-10 Geometria a informácie v stavebných procesoch: od komplexných systémov CAD a BIM- až po zjednodušené údaje pre analytiku.

Zatiaľ čo v IT boli neúspechy sémantického webu kompenzované nástupom nových technológií (big data, IoT, machine learning, AR/VR), stavebníctvo takéto príležitosti nemá.

Okrem výziev spojených s používaním pojmov na oznamovanie vzťahov medzi údajmi o prvkoch projektu zostáva základný problém - samotná dostupnosť týchto údajov. V stavebníctve stále prevládajú uzavreté systémy, čo sťaže prácu s údajmi, zdieľanie informácií a zvyšovanie efektívnosti procesov.

Práve uzavretosť údajov sa stáva jednou z hlavných prekážok, ktoré bránia rozvoju digitálnych riešení v stavebníctve. Na rozdiel od odvetvia IT, kde sa štandardom stali otvorené a harmonizované formáty údajov, v odvetví CAD (BIM) každý softvér používa svoj vlastný formát, čím sa vytvárajú uzavreté ekosystémy a umelo sa obmedzujú používateľia.



## KAPITOLA 6.2.

# UZAVRETÉ FORMÁTY PROJEKTOV A OTÁZKY INTEROPERABILITY

### **Uzavreté údaje a klesajúca produktivita: slepá ulička odvetvia CAD (BIM)**

Proprietárny charakter systémov CAD -systémov viedol k tomu, že každý program má svoj vlastný jedinečný formát údajov, ktorý je buď uzavretý a zvonku neprístupný - RVT, PLN, DWG, NDW, NWD, SKP, alebo je k dispozícii v pološtruktúrovanej forme prostredníctvom pomerne zložitého procesu konverzie JSON, XML (CPIXML), IFC, STEP a ifcXML, IfcJSON, BIMJSON, IfcSQL, CSV, atď..

Rôzne formáty údajov, v ktorých môžu byť uložené rovnaké údaje o rovnakých projektoch, sa líšia nielen štruktúrou, ale obsahujú aj rôzne verzie interného značenia, ktoré musia vývojári zohľadniť, aby zabezpečili kompatibilitu aplikácií. Napríklad formát CAD z roku 2025 sa otvorí v programe CAD z roku 2026, ale ten istý projekt sa nikdy neotvorí vo všetkých verziach programu CAD, ktoré mohli byť k dispozícii pred rokom 2025.

Tým, že poskytovateľ softvéru v stavebnictve neposkytuje priamy prístup k databázam, často vytvára svoj vlastný jedinečný formát a nástroje, ktoré musí odborník (projektant alebo správca údajov) používať na prístup, import a export údajov.

V dôsledku toho predajcovia základných CAD (BIM) a súvisiacich riešení (napr. ERP/PMIS)) neustále zvyšujú ceny za používanie produktov a bežní používateelia sú nútení platiť "províziu" v každej fáze prenosu údajov podľa formátov [63]: za pripojenie, import, export a prácu s údajmi, ktoré si používateelia sami vytvorili.

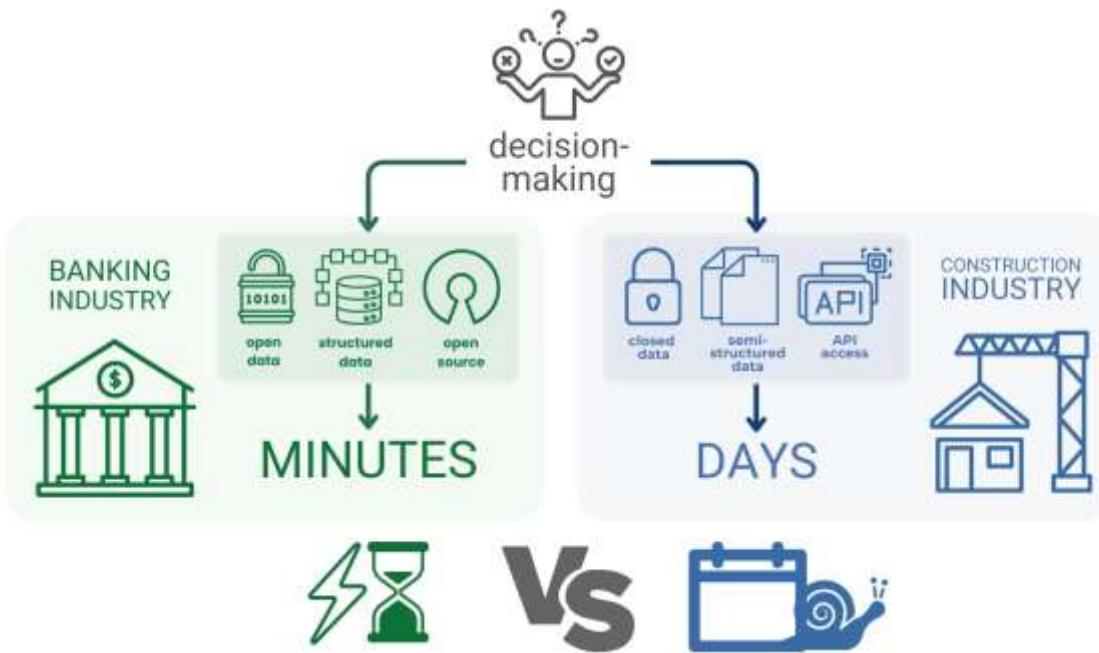
Náklady na prístup k údajom v clouдовom úložisku z populárnych produktov CAD - (BIM-) dosiahnu v roku 2025 1 dolár za transakciu [120] a predplatné stavebných ERP -produktov pre stredne veľké spoločnosti dosahuje päť- a šestciferné sumy ročne [121].

Podstata moderného stavebného softvéru spočíva v tom, že nie automatizácia alebo zvýšená efektívnosť, ale schopnosť inžinierov porozumieť konkrétnemu vysoko specializovanému softvéru ovplyvňuje kvalitu a náklady na spracovanie údajov o stavebných projektoch, ako aj zisky a dlhodobé prežitie spoločností realizujúcich stavebné projekty.

Nedostatočný prístup k databázam CAD -systémom, ktoré sa využívajú v desiatkach ďalších systémov a stovkách procesov [63], a z toho vyplývajúca nedostatočná kvalita komunikácie medzi jednotlivými odborníkmi priviedla stavebnictvo k statusu jedného z najneefektívnejších odvetví hospodárstva z hľadiska produktivity [44].

Za posledných 20 rokov projektových aplikácií CAD- (BIM-), vzniku nových systémov (ERP), nových stavebných technológií a materiálov klesla produktivita celého stavebnictva o 20% (obrázok 2.2-1), zatiaľ čo celková produktivita všetkých odvetví hospodárstva, ktoré nemajú veľké problémy s

priístupom k databázam a marketingovým koncepciam podobným BIM, sa zvýšila o 70% (vo výrobnom priemysle o 96%) [122].



Obr. 6.2-1 Vzhľadom na izolovanosť a zložitosť projektových údajov, od ktorých sú v stavebníctve závislé desiatky oddelení a stovky procesov, je rýchlosť rozhodovania niekoľkonásobne nižšia ako v iných odvetviach.

Existujú však aj ojedinelé príklady alternatívnych prístupov k vytváraniu interoperability medzi riešeniami CAD. Najväčšia európska stavebná spoločnosť s projektom SCOPE [123], ktorý sa začal ešte v roku 2018, ukazuje, ako je možné prekročiť klasickú logiku systémov CAD- (BIM-). Namiesto toho, aby sa snažili podriadiť IFC alebo sa spoliehali na proprietárne geometrické jadrá, vývojári SCOPE využívajú API a SDK reverzného inžinierstva na extrakciu údajov z rôznych CAD programov, ich konverziu do neutrálnych formátov, ako je OBJ alebo CPIXML, založených na jedinom open source geometrickom jadre OCCT, a ďalej ich aplikujú na stovky obchodných procesov stavebných a projektových spoločností. Napriek progresívnosti myšlienky však takéto projekty narážajú na obmedzenia a zložitosť slobodných geometrických jadier a stále súčasťou uzavretých ekosystémov jednej spoločnosti, ktoré reprodukujú logiku monovendorských riešení.

V dôsledku obmedzení uzavretých systémov a rozdielov v dátových formátoch, ako aj nedostatku účinných nástrojov na ich zjednotenie sa spoločnosti, ktoré musia pracovať s formátmi CAD, stretávajú s hromadením značného množstva údajov s rôzny stupňom štruktúry a uzavretosti. Tieto údaje sa nevyužívajú správne a miznú v archívoch, kde zostávajú navždy zabudnuté a nevyužité.

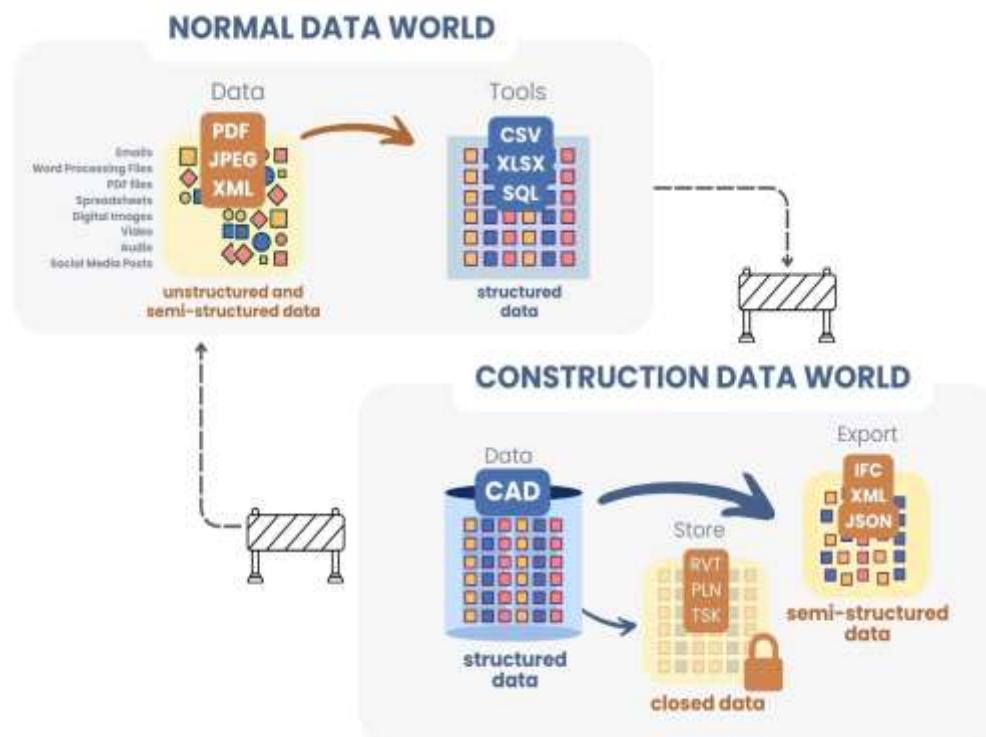
Údaje získané s veľkým úsilím vo fáze návrhu sa stávajú pre svoju zložitosť a uzavretosť nedostupnými na ďalšie použitie.

V dôsledku toho boli vývojári v stavebníctve za posledných 30 rokov nútení čeliť stále tomu istému

problému: každý nový uzavretý formát alebo proprietárne riešenie vyvoláva potrebu integrácie s existujúcimi otvorenými a uzavretými systémami CAD. Tieto neustále pokusy o zabezpečenie interoperability medzi rôznymi riešeniami CAD a BIM slúžia len na skomplikovanie dátového ekosystému namiesto toho, aby prispeli k jeho zjednodušeniu a štandardizácii.

## Mýtus o interoperabilite medzi systémami CAD

Zatiaľ čo v polovici 90. rokov bolo hlavným smerom vývoja interoperability v prostredí CAD prelomenie proprietárneho formátu DWG - čo vyvrcholilo víťazstvom aliancie Open DWG [75] a skutočným otvorením najpopulárnejšieho formátu výkresov pre celý stavebný priemysel - v polovici 20. rokov sa ľažisko presunulo. V stavebníctve naberá na sile nový trend: početné vývojové tímy sa zameriavajú na vytváranie takzvaných "mostov" medzi uzavretými systémami CAD (uzavretý BIM), formátom IFC a otvorenými riešeniami (otvorený BIM). Väčšina týchto iniciatív je založená na využívaní formátu IFC a geometrického jadra OCCT, ktoré poskytuje technický most medzi rozdielnymi platformami. Tento prístup sa považuje za slúbný smer, ktorý môže výrazne zlepšiť výmenu údajov a interoperabilitu softvérových nástrojov.

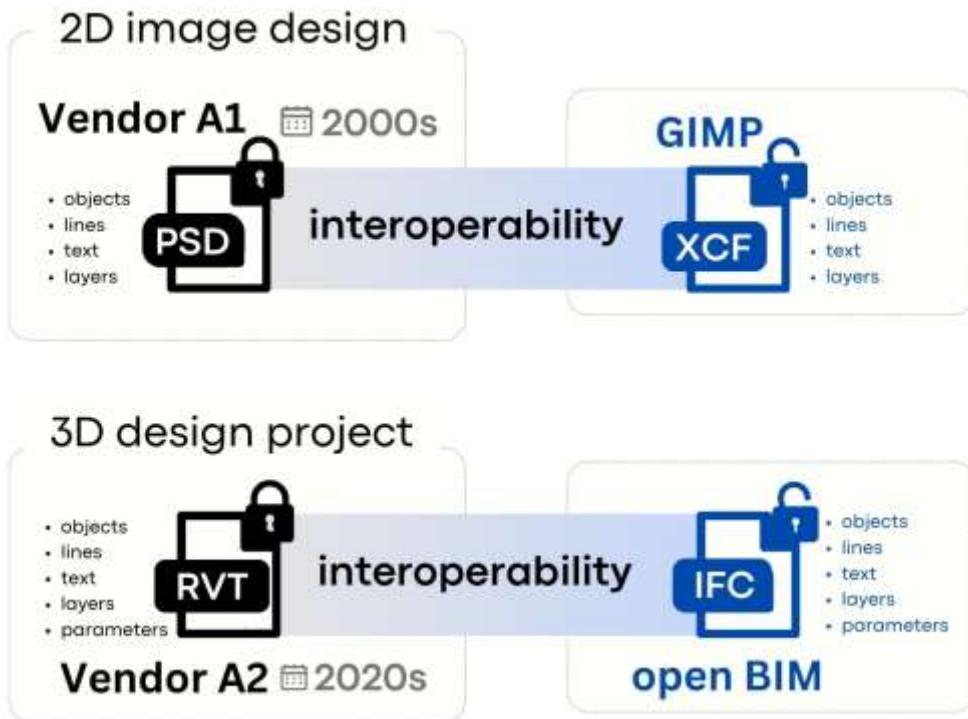


Obrázok 6.2-2 Zatiaľ čo iné odvetvia pracujú s otvorenými údajmi, stavebníctvo musí pracovať s uzavretými alebo voľne štruktúrovanými formátmi CAD (BIM).

Takýto prístup má historické paralely. V roku 2000 sa vývojári v snahe prekonáť dominanciu najväčšieho dodávateľa grafických editorov (2D world) pokúsili vytvoriť bezproblémovú integráciu medzi jeho proprietárnym riešením a slobodným open source - alternatívou ku GIMPu (obr. 6.2-3). Vtedy, rovnako ako dnes v konštrukcii, išlo o snahu premostiť uzavreté a otvorené systémy pri

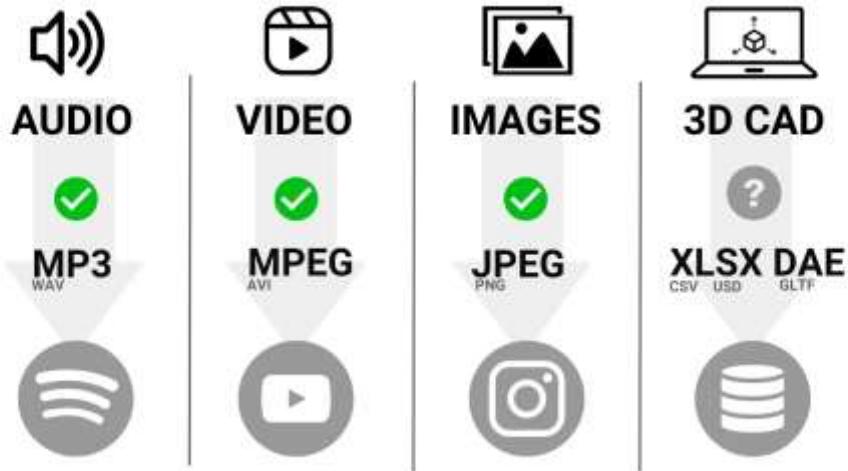
zachovaní komplexných parametrov, vrstiev a vnútornej logiky softvéru.

Používateľia však v skutočnosti hľadali jednoduché riešenia - ploché, otvorené údaje bez nadmernej zložitosti vrstiev a programových parametrov (obdoba geometrického jadra v CAD). Používateľia hľadali jednoduché a otvorené formáty údajov bez nadmernej logiky. Takýmito formátmi sa v grafike stali JPEG, PNG a GIF. Dnes sa používajú v sociálnych sieťach, na webových stránkach, v aplikáciách - sú ľahko spracovateľné a interpretovateľné bez ohľadu na platformu alebo výrobcu softvéru.



Obrázok 6.2-3 Interoperabilita dátových formátov v stavebníctve je podobná ceste od pokusov o zlúčenie proprietárneho produktu populárneho výrobcu a Open Source GIMP v roku 2000.

Výsledkom je, že takmer nikto v odvetví spracovania obrazu dnes nepoužíva uzavreté formáty ako PSD alebo otvorený XCF pre aplikácie, sociálne siete ako Facebook a Instagram alebo ako obsah na webových stránkach. Namiesto toho sa pri väčšine úloh používajú ploché a otvorené formáty JPEG, PNG a GIF z dôvodu jednoduchého používania a širokej kompatibility. Otvorené formáty ako JPEG a PNG sa stali štandardom pre zdieľanie obrázkov vďaka svojej univerzálnosti a širokej podpore, čo uľahčuje ich používanie na rôznych platformách. Podobný prechod možno pozorovať aj pri iných výmenných formátoch, ako sú video a audio, kde sa zdôrazňujú univerzálné formáty, ako sú MPEG a MP3, pre ich účinnosť kompresie a širokú kompatibilitu. Takýto prechod k štandardizácii zjednoduší zdieľanie a prehrávanie obsahu a informácií, vďaka čomu sú dostupné všetkým používateľom na rôznych platformách (obrázok 6.2-4).



Obrázok 6.2-4 Zjednodušené formáty bez zložitých funkcií úprav sa stali populárnymi na zdieľanie a používanie údajov.

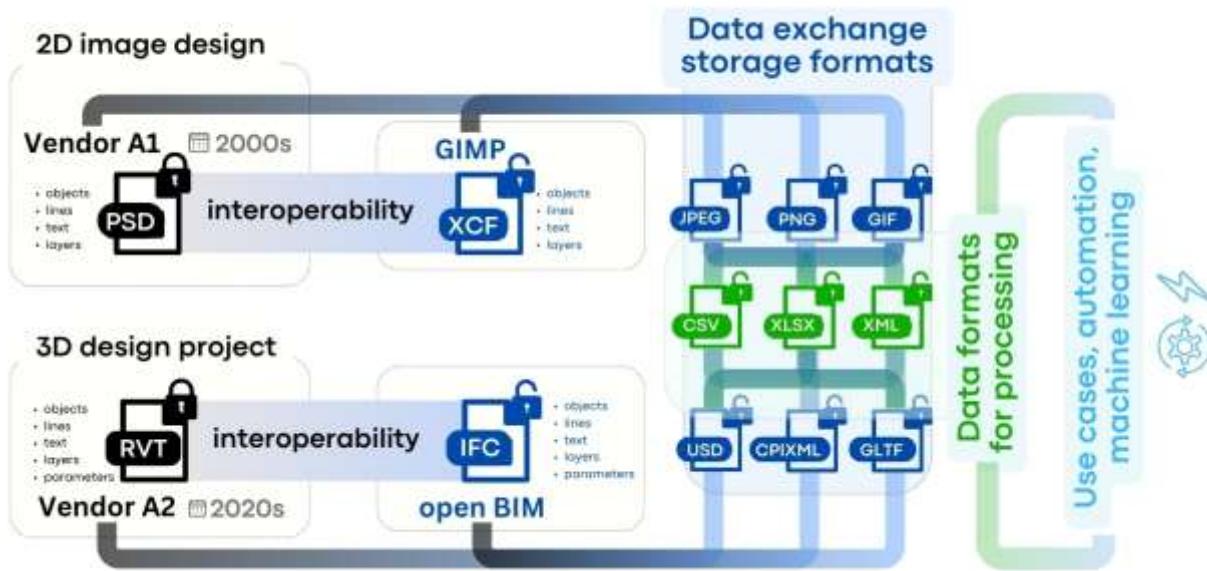
Podobné procesy prebiehajú aj pri 3D modelovaní. Jednoduché a otvorené formáty ako USD, OBJ, glTF, DAE, DXF, SQL a XLSX sa v projektoch čoraz viac používajú na výmenu údajov mimo prostredia CAD (BIM). Tieto formáty uchovávajú všetky potrebné informácie vrátane geometrie a metaúdajov bez potreby prevádzkovať zložitú štruktúru BREP, geometrické jadrá alebo interné klasifikátory špecifické pre dodávateľa. Proprietárne formáty, ako napríklad NWC, SVF, SVF2, CPIXML a CP2, ktoré poskytujú poprední dodávateľa softvéru, tiež plnia podobné funkcie, ale na rozdiel od otvorených štandardov zostávajú uzavreté.

Je pozoruhodné (a stojí za to znova pripomenúť, ako už bolo spomenuté v predchádzajúcej kapitole), že túto myšlienku - odmietnutie medzirodovo neutrálnych a parametrických formátov, ako sú IGES, STEP a IFC - podporil ešte v roku 2000 hlavný dodávateľ CAD, ktorý vytvoril Bielu knihu BIM a v roku 1994 zaregistroval formát IFC. V Bielej knihe z roku 2000 "Integrovaný návrh a výroba" [65] dodávateľ CAD zdôrazňuje dôležitosť natívneho prístupu k databáze CAD v rámci softvérového prostredia bez potreby používať medziľahlé prekladače a parametrické formáty, aby sa zachovala úplnosť a presnosť informácií.

Stavebníctvo sa zatiaľ nedohodlo ani na nástrojoch na prístup k databázam CAD, ani na ich nútenom reverznom inžinierstve, ani na priatí spoločného zjednodušeného formátu údajov na použitie mimo platforem CAD (BIM). Napríklad mnohé veľké spoločnosti v strednej Európe a nemecky hovoriačich regiónoch pôsobiace v stavebníctve používajú vo svojich ERP -systémoch formát CPIXML [121]. Tento proprietárny formát, ktorý je druhom XML, spája údaje CAD (BIM) projektu vrátane geometrických a metadát do jednej organizovanej zjednodušenej štruktúry. Veľké stavebné spoločnosti vytvárajú aj nové vlastné formáty a systémy, ako napríklad v projekte SCOPE, o ktorom sme hovorili v predchádzajúcej kapitole.

Uzavretá logika parametrických formátov CAD alebo komplexných parametrických súborov IFC (STEP) je vo väčšine obchodných procesov zbytočná. Používateľia hľadajú zjednodušené a ploché formáty,

ako sú USD, CPIXML, XML &OBJ, DXF, gITF, SQLite, DAE &XLSX, ktoré obsahujú všetky potrebné informácie o prvkoch, ale nie sú zaľažené nadbytočnou geometrickou logikou BREP, závislosťou od geometrických jadier a vnútorných klasifikácií konkrétnych produktov CAD a BIM -produktov (obr. 6.2-5).



Obr. 6.2-5 Pre väčšinu prípadov použitia si používateľia vyberajú najjednoduchšie možné formáty, ktoré sú nezávislé od programov dodávateľa.

Nástup plochých obrazových formátov, ako sú JPEG, PNG a GIF, oslobodených od nadbytočnej logiky interných mechanizmov výrobcov, podporil vývoj tisícov interoperabilných riešení na spracovanie a využívanie grafiky. To viedlo k vzniku aplikácií od nástrojov na retušovanie a filtrovanie až po sociálne siete, ako sú Instagram, Snapchat a Canva, kde sa tieto zjednodušené údaje dajú využiť bez toho, aby boli viazané na konkrétnego vývojára softvéru.

Štandardizácia a zjednodušenie projektových CAD-formátov podnieti vznik mnohých nových užívateľsky prívetivých a nezávislých nástrojov na prácu so stavebnými projektmi.

Odklon od zložitej logiky aplikácií dodávateľa viazaných na uzavreté geometrické jadrá k univerzálному otvoreným formátom založeným na knižničiach zjednodušených prvkov vytvára predpoklady na flexibilnejšie, transparentnejšie a efektívnejšie spracovanie údajov. Tým sa zároveň otvára prístup k informáciám pre všetky strany zapojené do procesu výstavby - od projektantov až po zákazníkov a servisné služby.

Napriek tomu je veľmi pravdepodobné, že v nasledujúcich rokoch sa dodávateľia CAD pokúsia opäť posunúť diskusiu o interoperabiliti a prístupe k databázam CAD. Pôjde už o "nové" koncepcie - ako sú granulárne údaje, inteligentné grafy, "federatívne modely", digitálne dvojčata v clouдовých úložiskách - ako aj o vytváranie priemyselných aliancií a štandardov, ktoré budú pokračovať v ceste BIM a otvoreného BIM. Napriek atraktívnej terminológii sa takéto iniciatívy môžu opäť stať nástrojmi

na udržanie používateľov v rámci proprietárnych ekosystémov. Jedným z príkladov je aktívna propagácia formátu USD (Universal Scene Description) ako "nového štandardu" pre spoluprácu medzi platformami CAD (BIM) od roku 2023.

## Prejsť na stránku USD a granulárne údaje m

Vznik aliancie AOUSD [124] v roku 2023 znamená dôležitý obrat v stavebníctve. Sme svedkami začiatku novej reality, ktorú formujú dodávateľia CAD, pri práci so stavebnými údajmi prostredníctvom niekoľkých významných zmien. Prvá významná zmena sa týka vnímania CAD -údajov. Odborníci, ktorí sa podieľajú na počiatočných fázach koncepčného návrhu, si čoraz viac uvedomujú, že vytvorenie návrhu v prostredí CAD je len východiskovým bodom. Údaje vytvorené počas procesu projektovania sa nakoniec stávajú základom pre analýzu, prevádzku a správu objektov. To znamená, že musia byť prístupné a použiteľné v systémoch nad rámec tradičných nástrojov CAD.

Zároveň dochádza k revolúcii v prístupe popredných vývojárov. Popredný dodávateľ CAD-, tvorca koncepcie BIM a formátu IFC, robí vo svojej stratégii nečakaný obrat. Od roku 2023 spoločnosť upúšťa od tradičného ukladania údajov v jednotlivých súboroch, zameriava sa na prácu s granulovanými (normalizovanými a štruktúrovanými) údajmi a prechádza na prístup zameraný na údaje [125].

Dodávateľia nasledujú historické trendy iných odvetví: väčšina používateľov nepotrebuje uzavreté formáty CAD (podobne ako PSD) alebo zložité parametrické súbory IFC (podobne ako GIMP s logikou vrstiev). Potrebujú jednoduché obrázky objektov, ktoré možno použiť v CAFM (stavebný Instagram), ERP (Facebook) a tisíce ďalších procesov naplnených tabuľkami Excel a dokumentmi PDF.

Súčasné trendy v stavebníctve potenciálne vytvárajú podmienky na postupný odklon od parametrických a zložitých formátov v prospech univerzálnejších a nezávislejších formátov USD, GLTF, DAE, OBJ (s metainformáciami v rámci hybridných aj samostatných štruktúrovaných alebo voľne štruktúrovaných formátov). Historickí lídri vrátane veľkých projekčných spoločností, ktoré kedysi v polovici 90. rokov aktívne presadzovali IFC, teraz otvorené propagujú nový formát USD [93] a zdôrazňujú jeho jednoduchosť a univerzálnosť (obr. 6.2-6). Masové prijatie USD v produktoch, kompatibilita s GLTF a aktívna integrácia do nástrojov, ako sú Blender, Unreal Engine a Omniverse, ukazujú potenciál začiatku novej paradigmy práce s údajmi. Spolu s popularitou lokalizovaných riešení, ako je európsky plochý formát USD - CPIXML, používaný v populárnych európskych ERP, by mohol potenciálne posilniť pozíciu USD v strednej Európe. Organizácie zapojené do vývoja formátu IFC už prispôsobujú svoju strategiu USD [126], čo len potvrdzuje nevyhnutnosť tohto posunu.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing.	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

Obrázok 6.2-6 Porovnanie technických špecifikácií formátov IFC a USD.

V tejto súvislosti má USD potenciál stať sa de facto štandardom a sľubuje prekonanie mnohých súčasných obmedzení, ktoré súvisia predovšetkým so zložitosťou existujúcich formátov CAD - (BIM-) a závislosťou ich interpretácie od geometrických jadier.

Namiesto parametrických a zložitých formátov CAD a IFC si v stavebníctve získajú miesto zjednodušené dátové formáty USD, gLTF, DAE, OBJ s metainformáciami o prvkoch v CSV, XLSX, JSON, XML vďaka svojej jednoduchosti a flexibilite.

Súčasné zmeny v stavebníctve na prvý pohľad vyzerajú ako technologický prelom spojený s prechodom zo zastaraného systému IFC na modernejší systém USD. Treba však vziať do úvahy, že

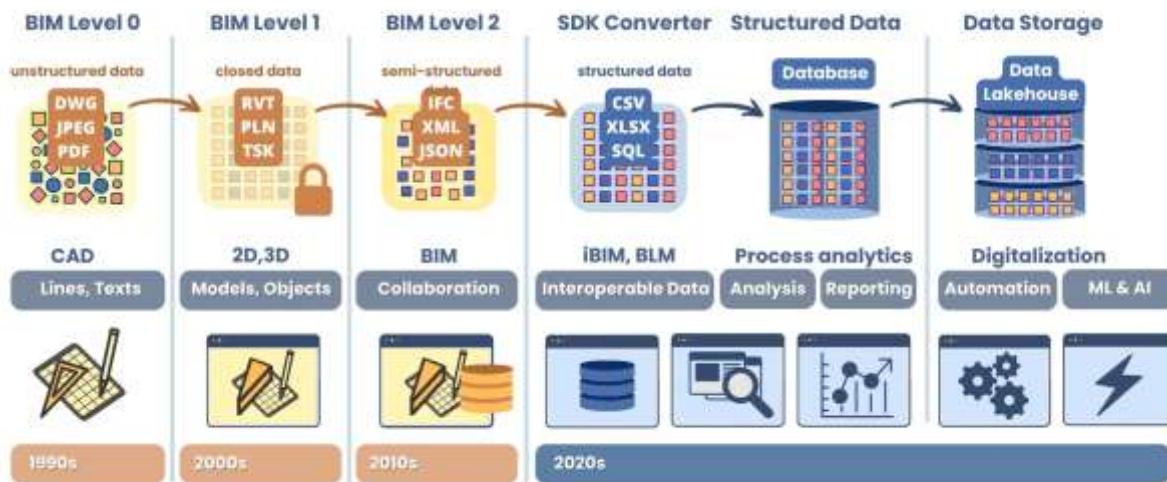
ešte v roku 2000 ten istý dodávateľ CAD, ktorý vyvinul IFC, písal o jeho problémoch a potrebe prístupu k databáze [65] a teraz aktívne propaguje prechod na nový štandard - USD.

Za ďalšou fasádou "otvorených údajov" USD a "nových" koncepcii granulárnej správy údajov prostredníctvom cloudových aplikácií, ktoré začínajú propagovať dodávateľia CAD, sa môže skrývať zámer dodávateľov monopolizovať správu projektových údajov, pričom používatelia sa ocitnú v situácii, keď výber formátu súvisí viac s podnikovými záujmami ako so skutočnými potrebami.

Analýza kľúčových skutočností [93] ukazuje, že hlavným cieľom týchto zmien nie je ani tak pohodlie používateľov, ako predovšetkým zachovanie kontroly nad ekosystémami a dátovými tokmi v prospech dodávateľov, ktorí za 40 rokov nikdy neboli schopní poskytnúť prístup k databázam CAD.

Možno teraz nastal čas, aby spoločnosti prestali čakať na nové koncepty od dodávateľov softvéru a zamerali sa na vlastný vývoj v dátovo orientovanom smere. Po osloboodení sa od problémov s prístupom k údajom prostredníctvom nástrojov reverzného inžinierstva bude priemysel schopný samostatne prejsť na moderné, bezplatné a pohodlné nástroje na prácu a analýzu údajov bez toho, aby bolo potrebné vnucovať nové koncepcie.

### CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Obr. 6.2-7 Úroveň vyspelosti CAD (BIM): od neštruktúrovaných údajov k štruktúrovaným údajom a úložiskám.

Prístup k databázam, otvorené údaje a formáty sa nevyhnutne stanú štandardom v stavebníctve bez ohľadu na pokusy dodávateľov tento proces zdržať - je to len otázka času (obr. 6.2-7). Tempo tohto prechodu sa môže výrazne zvýšiť, ak sa čoraz viac odborníkov zoznámi s otvorenými formátmi, databázovými nástrojmi a dostupnými SDK pre reverzné inžinierstvo, ktoré umožňujú priamy prístup k údajom CAD -systémov [92].

Budúcnosť je v otvorených, jednotných a analyticky dostupných údajoch. Aby sa stavebné a inžinierske spoločnosti vyhli závislosti od dodávateľských riešení a nestali sa rukojemníkmi uzavretých ekosystémov, budú sa musieť skôr či neskôr spoľahnúť na otvorenosť a nezávislosť a vybrať si formáty a riešenia, ktoré poskytujú úplnú kontrolu nad údajmi.

Údaje, ktoré sa dnes vytvárajú v stavebníctve, budú v budúcnosti kľúčovým zdrojom pre obchodné rozhodnutia. Budú pôsobiť ako strategické "palivo", ktoré poháňa rozvoj a efektívnosť stavebných spoločností. Budúcnosť stavebníctva spočíva v schopnosti pracovať s údajmi, nie vo výbere dátových formátov alebo modelov.

Na pochopenie rozdielu medzi otvorenými formátmi USD, glTF, DAE, OBJ a proprietárnymi parametrickými formátmi CAD je dôležité vziať do úvahy jeden z najzložitejších a kľúčových dátových prvkov pri vizualizácii a výpočtoch návrhu - geometriu a procesy jej generovania. A aby sme pochopili, ako sa geometrické údaje stávajú základom pre analytické a výpočty v stavebníctve, je potrebné sa hlbšie pozrieť na mechanizmy generovania, transformácie a ukladania geometrie.



## KAPITOLA 6.3.

### GEOMETRIA V STAVEBNÍCTVE: OD ČIAR PO METRE KUBICKÉ

#### **Ked' sa čiary menia na peniaze alebo prečo stavitelia potrebujú geometriu**

Geometria v stavebníctve nie je len vizualizáciou, ale aj základom presných kvantitatívnych výpočtov. V modeli projektu geometria dopĺňa zoznamy parametrov prvkov (obr. 3.1-16) o dôležité objemové charakteristiky, ako sú dĺžka, plocha a objem. Hodnoty týchto objemových parametrov sa počítajú automaticky pomocou jadier geometrie a sú východiskom pre odhady, harmonogramy a modely zdrojov. Ako sme už uviedli v 5. časti tejto knihy a v kapitole "Kalkulácia a odhadovanie nákladov stavebných projektov", práve objemové parametre skupín objektov z CAD -modelov tvoria základ moderných ERP, PMIS -systémov/ Geometria zohráva zásadnú úlohu nielen vo fáze projektovania, ale aj v riadení realizácie projektu, kontrole harmonogramu, rozpočtovaní a prevádzke. Tak ako pred tisícami rokov pri stavbe egyptských pyramíd závisela presnosť projektu od dĺžkových mier, ako sú lakte a kubity, dnes presnosť interpretácie geometrie v CAD -programoch priamo ovplyvňuje výsledok: od rozpočtu a termínov až po výber dodávateľa a logistiku dodávok

Vo vysoko konkurenčnom prostredí s obmedzeným rozpočtom sa presnosť objemových výpočtov, ktorá priamo závisí od geometrie, stáva faktorom prežitia. Moderné ERP - systémy priamo závisia od správnych objemových charakteristík získaných z CAD - a BIM -modelov. Preto presný geometrický opis prvkov nie je len vizualizáciou, ale kľúčovým nástrojom na riadenie nákladov a času výstavby.

Geometria bola v minulosti hlavným jazykom inžinierskej komunikácie. Od čiar na papyrusu až po digitálne modely slúžili výkresy a geometrické znázornenia ako prostriedok výmeny informácií medzi projektantmi, superintendentmi a odhadcami. Pred príchodom počítačov sa výpočty vykonávali ručne pomocou pravítok a uhlomerov. Dnes je táto úloha automatizovaná vďaka objemovému modelovaniu: geometrické jadrá softvéru CAD premieňajú čiary a body na trojrozmerné telesá, z ktorých sa automaticky získavajú všetky potrebné charakteristiky.

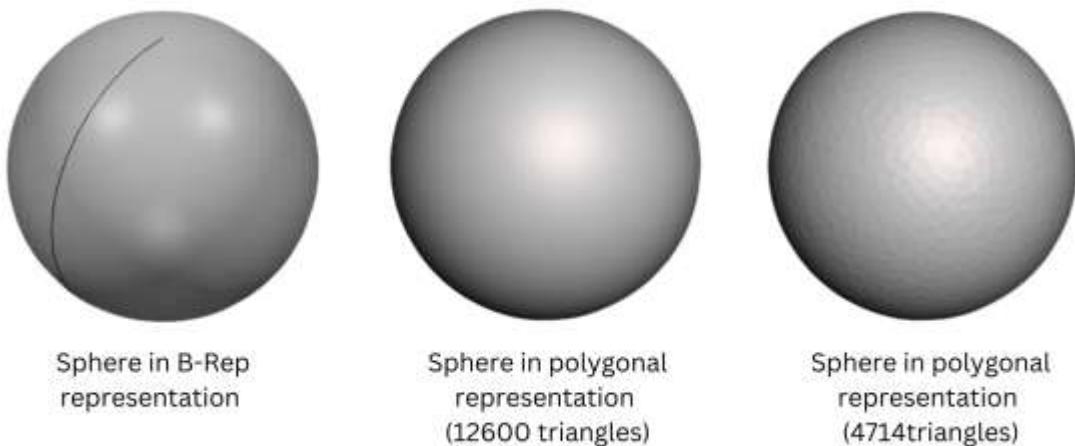
Pri práci v CAD -programoch sa vytváranie geometrických prvkov pre výpočty vykonáva prostredníctvom používateľského rozhrania CAD- (BIM-) programov. Na transformáciu bodov a čiar na objemové telesá sa používa geometrické jadro, ktoré plní kľúčovú úlohu - transformáciu geometrie na objemové modely, z ktorých sa po aproximácii automaticky vypočítajú objemové charakteristiky prvku.

#### **Od čiar k objemom: Ako sa z plochy a objemu stávajú údaje**

V inžinierskej praxi sa objemy a plochy počítajú z geometrických plôch opísaných analyticky alebo prostredníctvom parametrických modelov, ako sú NURBS (nonuniform rational B-splines) v rámci BREP (boundary element representation).

NURBS (Non-Uniform Rational B-Splines) je matematický spôsob opisu krviek a povrchov, zatiaľ čo BREP je rámec na opis kompletnej trojrozmernej geometrie objektu vrátane jeho hraníc, ktoré možno definovať pomocou NURBS.

Napriek presnosti BREP a NURBS si vyžadujú výkonné výpočtové zdroje a zložité algoritmy. Priamy výpočet z takýchto matematicky presných opisov je však často výpočtovo náročný, preto sa v praxi takmer vždy používa teselácia - transformácia povrchov na sietť trojuholníkov -, ktorá zjednodušuje následné výpočty. Teselácia je rozdelenie komplexného povrchu na trojuholníky alebo mnogouholníky. V prostrediach CAD /CAE sa táto metóda používa na vizualizáciu, objemové výpočty, vyhľadávanie kolízií, export do formátov ako MESH a analýzu kolízií. Príkladom z prírody je včelia plást, kde sa zložitý tvar rozkladá na pravidelnú sietť (obr. 6.3-1).



Obr. 6.3-1 Tá istá guľa v parametrickom popise BREP a polygonálnom zobrazení rôznym počtom trojuholníkov.

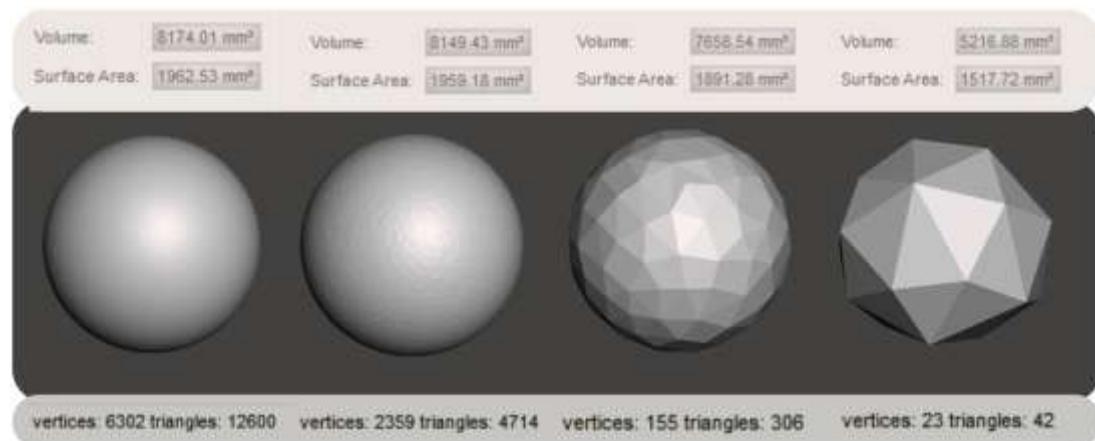
BREP (NURBS), ktorý sa používa v CAD, nie je základným modelom geometrie. Bol vytvorený ako vhodný nástroj na reprezentáciu kružníc a racionálnych splajnov a na minimalizáciu ukladania geometrických údajov. Má však obmedzenia - napríklad neschopnosť presne opísť sínusoidu, ktorá je základom špirálových čiar a plôch, a potrebu používať zložité geometrické jadrá.

Naopak, trojuholníkové siete a teselácia parametrických tvarov sa vyznačujú jednoduchosťou, efektívnym využívaním pamäte a schopnosťou spracovať veľké množstvo údajov (obr. 6.3-2). Tieto výhody umožňujú zaobísť sa pri výpočte geometrických tvarov bez zložitých a drahých geometrických jadier, a v nich obsiahnutých desiatok miliónov riadkov kódu.

Vo väčšine prípadov budov nezáleží na tom, ako presne sú objemové charakteristiky definované - prostredníctvom parametrických modelov (BREP, IFC) alebo prostredníctvom polygónov (USD, glTF, DAE, OBJ). Geometria zostáva formou aproximácie: či už prostredníctvom NURBS alebo MESH, vždy ide o približný opis tvaru.

Geometria definovaná ako polygóny alebo BREP (NURBS) zostáva do určitej miery len spôsobom aproximácie s približným opisom spojitého tvaru. Tak ako Fresnelove integrály nemajú presné analytické vyjadrenie, diskretizácia geometrie prostredníctvom polygónov alebo NURBS je vždy aproximáciou, rovnako ako trojuholníkový MESH.

Parametrická geometria vo formáte BREP je potrebná hlavne tam, kde je dôležitá minimálna veľkosť údajov a na jej spracovanie a zobrazenie je možné použiť geometrické jadrá náročné na zdroje a drahé. Najčastejšie je to charakteristické pre vývojárov CAD -programov, ktorí na tento účel používajú vo svojich produktoch geometrické jadrá MCAD -dodávateľov. V tomto prípade sa aj v rámci týchto programov BREP-modely v procese teselácie na vizualizáciu a výpočty často konvertujú na trojuholníky (podobne ako sa PSD-súbory zjednodušujú na JPEG).



Obr. 6.3-2 Rozdiel objemových charakteristik v obrázkoch s rôznym počtom polygónov.

Polygonálny MESH, ako aj parametrický BREP, majú svoje výhody a obmedzenia, ale cieľ je rovnaký - opísť geometriu s prihľadnutím na úlohy používateľa. V konečnom dôsledku presnosť geometrického modelu závisí nielen od metódy jeho reprezentácie, ale aj od požiadaviek konkrétnej úlohy.

Vo väčšine konštrukčných problémov môže byť potreba parametrickej geometrie a zložitých geometrických jadier zbytočná.

Pri každej konkrétnej úlohe automatizácie výpočtu stojí za to zvážiť, či význam parametrickej geometrie nie je preceňovaný vývojármami CAD, ktorí majú záujem propagovať a predávať svoje vlastné softvérové produkty.

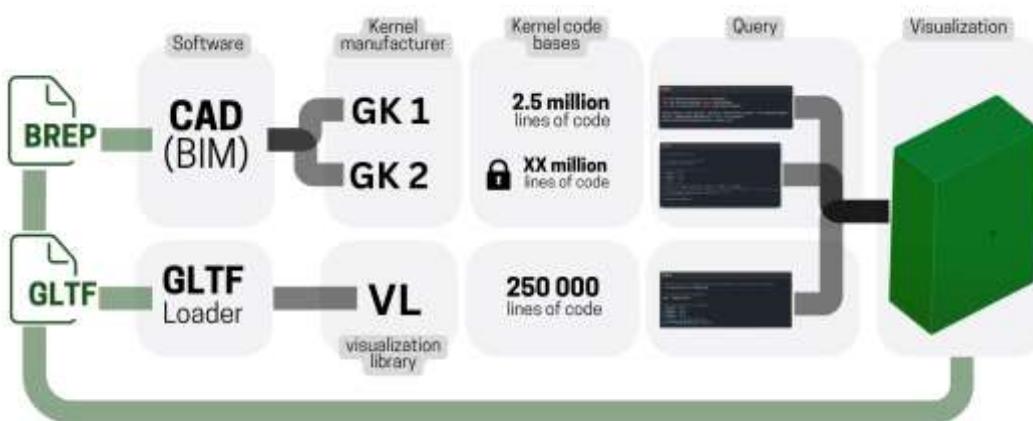
## Prechod na MESH, USD a polygóny: používanie teselačného zobrazenia pre geometriu

V stavebníctve je pri streamovaní, vývoji systémov, databáz alebo automatizácií procesov, ktoré pracujú s konštrukčnými informáciami a geometriou prvkov, dôležité usilovať sa o nezávislosť od konkrétnych CAD editorov a geometrických jadier.

Výmenný formát, ktorý sa má používať vo výpočtových oddeleniach a na stavenisku, by nemal byť založený na konkrétnom programe CAD- (BIM-). Geometrické informácie by mali byť vo formáte reprezentované priamo prostredníctvom teselácie bez odkazu na geometrické jadro alebo CAD architektúru.

Parametrickú geometriu z CAD možno považovať za medzistupeň, ale nie za základ univerzálneho formátu. Väčšina parametrických opisov (vrátane BREP a NURBS) sa v každom prípade pre ďalšie spracovanie konvertuje na polygonálny MESH. Ak je výsledok rovnaký (teselácia a polygóny) a proces je jednoduchší, volba je zrejmá. Je to analogické s volbou medzi grafovými ontológiami a štruktúrovanými tabuľkami (o ktorých sme hovorili v štvrtej časti): prílišná zložitosť je zriedkakedy opodstatnená (obr. 3.2-10, obr. 6.1-8).

Otvorené formáty, ako napríklad: OBJ, STL, glTF, SVF, CPIXML, USD a DAE, využívajú univerzálnu štruktúru trojuholníkovej siete, čo im poskytuje značné výhody. Tieto formáty majú vynikajúcu interoperabilitu - dajú sa ľahko čítať a vizualizovať pomocou dostupných knižníc s otvoreným zdrojovým kódom bez potreby zložitých špecializovaných geometrických jadier obsahujúcich milióny riadkov kódu (obrázok 6.3-3). Tieto univerzálné geometrické formáty sa používajú v aplikáciách od relatívne jednoduchých nástrojov na navrhovanie kuchýň v IKEA™ až po komplexné systémy vizualizácie objektov v kine a VR-aplikáciách. Dôležitou výhodou je dostupnosť veľkého množstva bezplatných a otvorených knižníc na prácu s týmito formátmi, ktoré sú k dispozícii pre väčšinu platform a programovacích jazykov.



Obrázok 6.3-3 Rovnakú reprezentáciu geometrie možno dosiahnuť pomocou parametrických formátov a geometrických jadier, alebo pomocou triangulačných formátov a knižníc pre vizualizáciu s otvoreným zdrojovým kódom.

Rovnako ako samotní používateľia, aj dodávatelia CAD majú problémy s interpretáciou zahraničných parametrických CAD formátov alebo otvorených IFC kvôli odlišným geometrickým jadrám. V praxi všetci dodávatelia CAD -predajcovia bez výnimky používajú na prenos údajov medzi systémami reverzné inžinierstvo SDK a žiadny z nich sa na účely interoperability nespolieha na formáty ako IFC alebo USD [93].

Namiesto používania konceptov propagovaných alianciami dodávateľov CAD-, ktoré sami nepoužívajú, je pre vývojárov a používateľov CAD riešení produktívnejšie zamierať sa na pochopenie výhod každého prístupu v konkrétnom kontexte a vybrať si jeden alebo druhý typ geometrie v závislosti od prípadu použitia. Výber medzi rôznymi geometrickými reprezentáciami je kompromisom medzi presnosťou, výpočtovou efektívnošťou a praktickými potrebami konkrétnej úlohy.

Zložitosť spojená s používaním geometrických jadier, ktorú tradične vnucujú stavebníctvu veľkí dodávatelia pri spracovaní projektových údajov, sa často ukazuje ako zbytočná. Formát USD založený na geometrii MESH sa môže stať pre priemysel akousi "Pandorinou skrinkou", ktorá vývojárom otvára nové možnosti organizácie výmeny údajov - mimo rámca štruktúr IFC a parametrických štruktúr BREP typických pre dodávateľov CAD.

Po bližšom pohľade na štruktúru formátov USD, DAE, gLTF, OBJ atď. je zrejmé, že existujú jednoduchšie, otvorené formáty, ktoré umožňujú efektívne organizovať prenos a používanie geometrických informácií bez potreby spoliehať sa na zložité parametrické a uzavreté geometrické jadrá. Tento prístup nielenže znižuje technický prah vstupu pre vývojárov, ale tiež podporuje vývoj flexibilných, škálovateľných a skutočne otvorených riešení pre digitálnu výstavbu.

### **LOD, LOI, LOMD - jedinečná klasifikácia detailov v CAD (BIM).**

Okrem formátov geometrickej reprezentácie vo svete, kde rôzne odvetvia používajú rôzne úrovne podrobnosti a hĺbky údajov, ponúkajú metodiky CAD - (BIM-) vlastné jedinečné klasifikačné systémy, ktoré štruktúrujú prístup k informovaniu modelov budov.

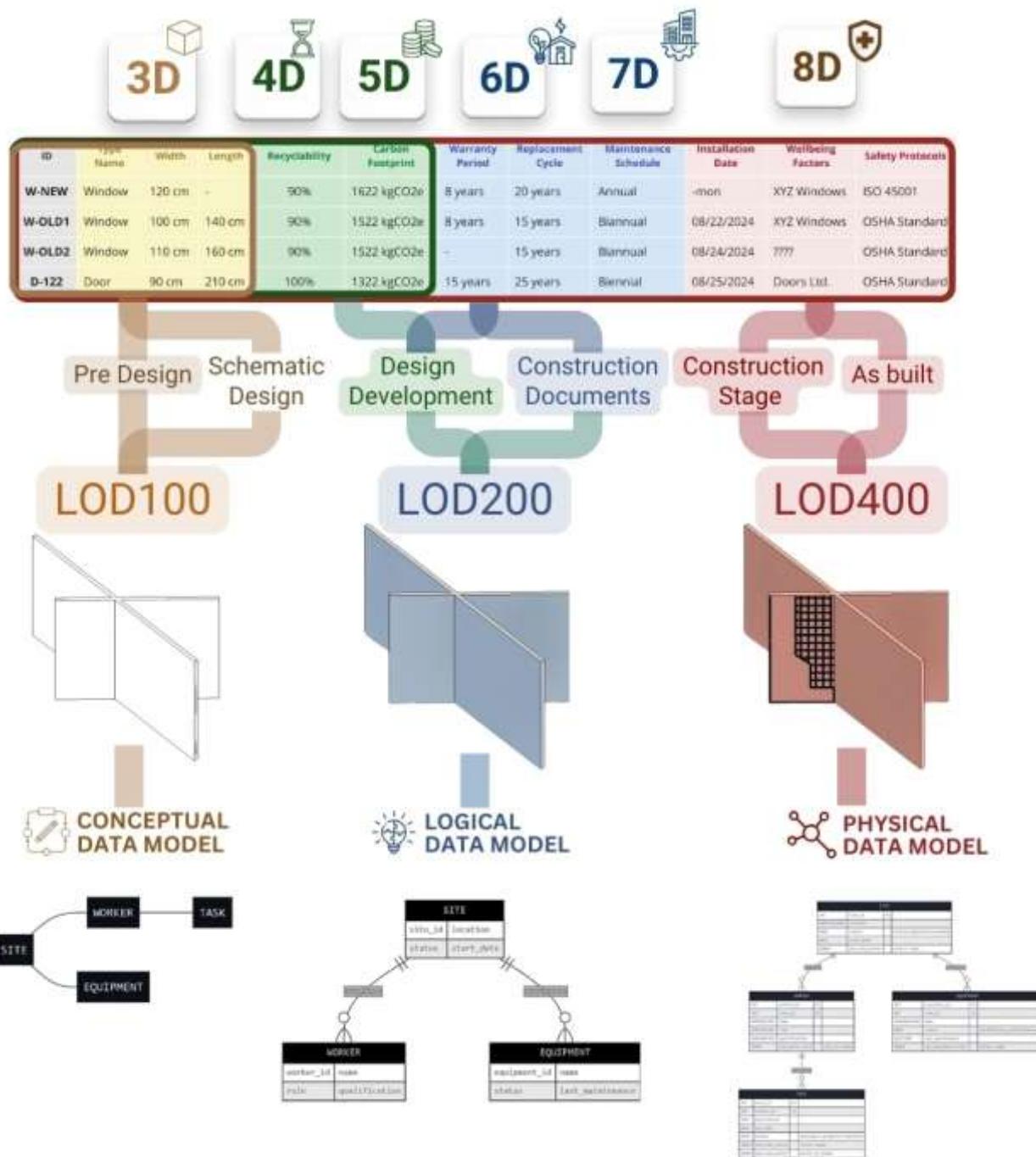
Jedným z príkladov nových prístupov k štandardizácii je zavedenie úrovní vývoja modelu, ktoré odrážajú stupeň pripravenosti a spoľahlivosti grafických aj informačných komponentov. Na rozlíšenie informačného obsahu pri práci s údajmi CAD - (BIM-) sa objavili LOD (Level Of Detail) - úroveň podrobnosti grafickej časti modelu a LOI (Level Of Information) - úroveň rozpracovanosti údajov. Okrem toho sa pre integrovaný prístup zaviedol pojem LOA (Level of Accuracy) - presnosť prezentovaných prvkov a LOG (Level of Geometry) na určenie presnosti grafického zobrazenia.

Úrovne podrobnosti (LOD) sú označené číslami od 100 do 500, ktoré odrážajú stupeň rozpracovanosti modelu. LOD 100 je koncepcný model so všeobecnými tvarmi a rozmermi. LOD 200 obsahuje presnejšie rozmery a tvary, ale s podmienenou podrobnosťou. LOD 300 je podrobný model s presnými rozmermi, tvarmi a umiestnením prvkov. LOD 400 obsahuje podrobné informácie potrebné na výrobu a inštaláciu prvkov. LOD 500 odráža skutočný stav zariadenia po výstavbe a používa sa na prevádzku a údržbu. Tieto úrovne opisujú štruktúru nasýtenia modelu CAD (BIM) informáciami v rôznych fázach životného cyklu vrátane 3D, 4D, 5D a ďalších.

V reálnych projektoch je vysoká úroveň podrobnosti (LOD400) často nadmerná a stačí použiť geometriu LOD100 alebo dokonca ploché výkresy, zatiaľ čo ostatné údaje možno získať buď výpočtom, alebo zo súvisiacich prvkov, ktoré nemusia mať odlišnú geometriu. Napríklad priestory a prvky miestností (kategórie prvkov miestností) nemusia mať vizuálnu geometriu, ale obsahujú značné množstvo informácií a databáz, okolo ktorých je postavených mnoho obchodných procesov.

Preto je dôležité jasne definovať požadovanú úroveň podrobnosti ešte pred začatím návrhu. Pre prípady použitia 4D -7D často postačujú aj výkresy DWG a minimálna geometria LOD100. Klúčovou úlohou v procese tvorby požiadaviek je nájsť rovnováhu medzi bohatosťou a praktickosťou modelu.

V podstate, ak považujeme údaje CAD (BIM) za databázu (ktorou aj sú), opis nasýtenia modelu prostredníctvom nových skratiek nie je ničím iným ako postupným modelovaním údajov pre informačné systémy, počnúc konceptuálnou úrovňou a končiac fyzickou (obr. 6.3-4), ktorému sme sa podrobne venovali v tretej a štvrtnej časti knihy. Každé zvýšenie LOD a LOI znamená pridanie informácií potrebných pre nové úlohy: výpočty, riadenie stavby, prevádzku a je charakterizované postupným obohacovaním modelu o ďalšie informačné vrstvy (3D -8D) v podobe rôznych parametrov, ktorým sme sa venovali v piatej časti knihy.



**Obrázok 6.3-4 Proces spresňovania detailov projektu je totožný s modelovaním údajov od koncepcného po fyzický model údajov.**

Geometria je len časťou konštrukčných údajov, ktorých potreba nie je v stavebných projektoch vždy opodstatnená a kľúčovým problémom práce s CAD -dátami nie je ani tak spôsob vizualizácie modelov, ale skôr to, ako sa dajú údaje z týchto modelov použiť mimo CAD- (BIM-) programov.

V polovici roku 2000 čelil stavebný priemysel bezprecedentnej výzve v súvislosti s rýchlym nárastom množstva údajov v systémoch riadenia a spracovania údajov, najmä tých, ktoré pochádzajú z oddelení

CAD (BIM). Tento dramatický nárast objemu údajov zaskočil manažérov spoločností, ktorí neboli pripravení na rastúce požiadavky na kvalitu a správu údajov.

## Nové normy CAD (BIM) - AIA, BEP, IDS, LOD, COBie

Využívajúc nedostatok otvoreného prístupu k databázam CAD a obmedzenú konkurenciu na trhu spracovania údajov a využívajúc marketingové kampane spojené s novou skratkou BIM, organizácie zaoberejúce sa vývojom prístupov k práci s údajmi CAD - začali vytvárať nové normy a koncepcie, ktoré by de iure mali byť zamerané na zlepšenie postupov správy údajov.

Hoci takmer všetky iniciatívy priamo alebo nepriamo podporované dodávateľmi CAD a vývojármami (BIM) boli zamerané na optimalizáciu pracovných postupov, vyústili do množstva nariem, za ktoré lobovali rôzne zainteresované strany, čo viedlo v stavebníctve k určitej nejednoznačnosti a zmätku v oblasti dátových procesov.

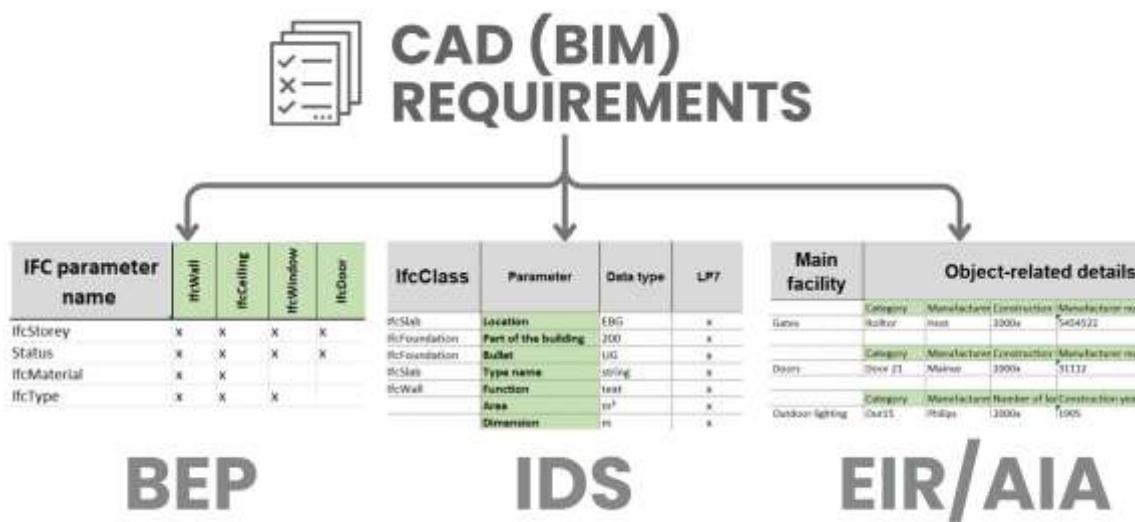
Uvedme si niektoré nové dátové štandardy, okrem LOD, LOI, LOA, LOG, ktoré sa v posledných rokoch objavili v stavebníctve:

- **BEP** (BIM Execution Plan) - popisuje, ako integrovať a využívať CAD (BIM) v projekte, definuje metódy a procesy spracovania údajov.
- **Dokument EIR /AIA** (Požiadavky na informácie objednávateľa) - pripravuje ho objednávateľ pred vyhlásením verejného obstarávania a obsahuje požiadavky na zhotoviteľa na prípravu a poskytnutie informácií. Slúži ako podklad pre BEP v príslušnom projekte.
- **AIM** (Asset Information Model) je súčasťou procesu BIM. Po odovzdaní a dokončení projektu sa dátový model nazýva informačný model aktíva alebo AIM. Účelom AIM je správa, údržba a prevádzka realizovaného aktíva.
- **IDS** (Information Delivery Specification) - definuje požiadavky a to, aké údaje a v akom formáte sa vyžadujú v rôznych fázach stavebného projektu.
- **iLOD** je úroveň podrobnosti LOD, s ktorou sú informácie zastúpené v modeli BIM. Definuje, aké podrobné a úplné sú informácie v modeli, od základných geometrických zobrazení až po podrobnejšie špecifikácie a údaje.
- **eLOD** - LOD úroveň podrobnosti jednotlivých prvkov v modeli CAD (BIM). Definuje rozsah, v akom je každý prvek modelovaný, a súvisiace informácie, ako sú rozmery, materiály, výkonnostné charakteristiky a ďalšie relevantné atribúty.
- **APS** (Platform Services) a ďalšie produkty od hlavných dodávateľov CAD (BIM) - popisujú nástroje a infraštruktúru potrebnú na vytváranie prepojených a otvorených dátových modelov.

Hoci deklarovaným cieľom zavádzania nariem CAD (BIM) - ako sú LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - je zlepšiť kvalitu správy údajov a rozšíriť možnosti automatizácie, v praxi ich používanie často vedie k nadmernej zložitosti a roztrieštenosti procesov. Ak považujeme model CAD (BIM) za druh databázy, je zrejmé, že mnohé z týchto nariem duplikujú dlhodobo zavedené a efektívne

priistupy používané v iných odvetviach pri práci s informačnými systémami. Namiesto zjednodušenia a zjednotenia takéto iniciatívy často vytvárajú ďalšiu terminologickú záťaž a bránia implementácii skutočne otvorených a flexibilných riešení.

Je pozoruhodné, že mnohé z týchto nových koncepcí v skutočnosti nahradzajú procesy modelovania a overovania údajov, ktoré boli podrobne rozoberané v prvých častiach knihy a ktoré sa už dlho používajú v iných odvetviach hospodárstva. Na druhej strane v stavebníctve sa proces štandardizácie často ubera opačným smerom - vznikajú nové formáty opisu údajov, nové normy a nové koncepcie validácie údajov, ktoré nie vždy vedú k skutočnej jednotnosti a praktickej použiteľnosti. Výsledkom je, že namiesto zjednodušenia a automatizácie spracovania sa odvetvie stretáva s ďalšími úrovňami regulácie a byrokracie (obr. 6.3-1), čo nie vždy prispieva k zvýšeniu efektívnosti.



Obr. 6.3-1 Požiadavky na obsah údajov a informácií sú redukované na opis atribútov a ich hraničných hodnôt, ktoré sú opísané pomocou tabuľiek.

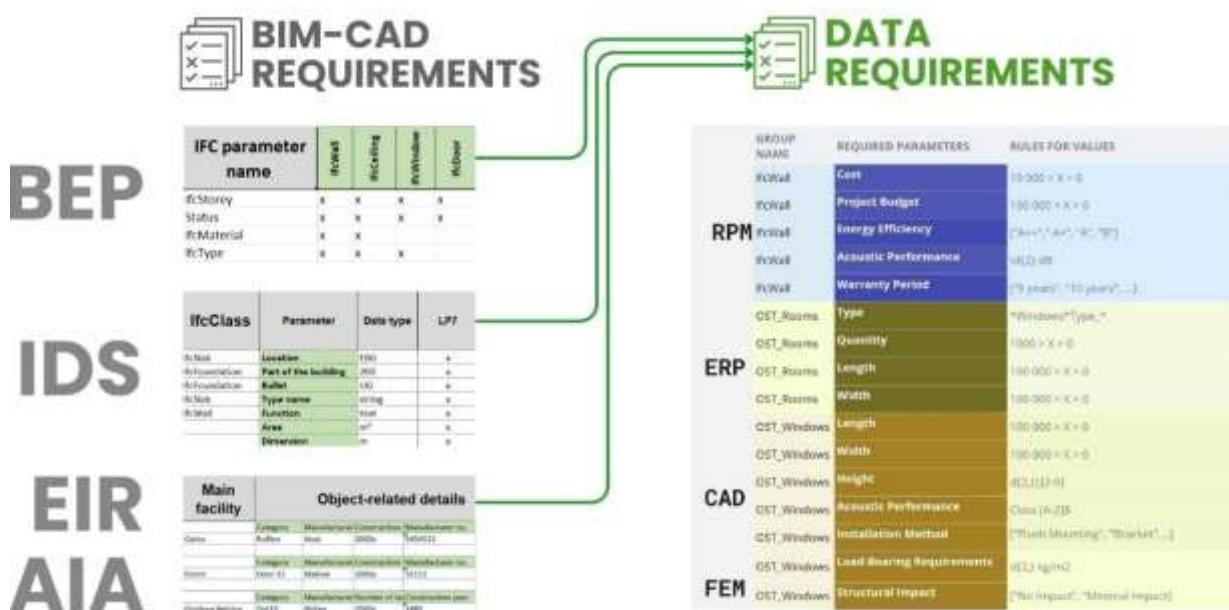
Namiesto zjednodušenia spracovania údajov nové koncepcie súvisiace s údajmi CAD (BIM) najčastejšie vytvárajú ďalšie zložitosti a spory už vo fáze interpretácie a základných definícií.

Jedným z najnovších príkladov nových koncepcí je formát IDS (zavedený v roku 2020), ktorý umožňuje opísť požiadavky na zloženie atribútov informačného modelu v otvorenej koncepcii BIM. Požiadavky IDS opisujú informácie o atribútoch a ich hraničných hodnotách vo forme štruktúrovanej tabuľky (Excel alebo MySQL), ktorá sa potom prekladá do značiek pološtruktúrovaného formátu XML, premenovaného z XML na špeciálnu skratku IDS.

Na rozdiel od názoru presadzovaného dodávateľmi a podporovaného BIM a otvoreným BIM, že spracovanie údajov v stavebníctve je jedinečné z dôvodu používania špecializovaných nástrojov, ako sú CAD a BIM, formáty údajov a postupy správy údajov v tomto odvetví sa nelisia od postupov iných odvetviach.

Počet požiadaviek na projekty a formáty CAD (BIM) možno zjednodušiť použitím jednej tabuľky požiadaviek s atribútmi-stípcami, podrobne opisanými v kapitole "Preklad požiadaviek do štruktúrovanej formy", bez toho, aby bolo potrebné pôvodne štruktúrované požiadavky preklaďať do netabuľkových formátov (IDS je pôvodne opísaný prostredníctvom tabuľky).

Zjednodušený prístup (obrázok 6.3-2), ktorý obsahuje stĺpce pre identifikátory entít, vlastnosti a hraničné hodnoty, ktoré boli podrobne rozobrané v predchádzajúcich kapitolách (obrázok 4.4-9, obrázok 4.4-16, obrázok 7.3-10), eliminuje potrebu konvertovať požiadavky do formátu IDS-XML. Táto metóda poskytuje priamy, menej ťažkopádny a transparentnejší mechanizmus kontroly kvality údajov. Spolieha sa na široko používané nástroje, od regulárnych výrazov (RegEx) po dátové rámce, Pandas a štandardné ETL -payplanes - presne také, aké používajú odborníci v iných odvetviach hospodárstva na prácu s údajmi.



Obr. 6.3-2 Požiadavky na údaje v iných odvetviach sú zjednodušené na štruktúrovaný opis atribútov a ich hraničných hodnôt.

Postupom času sa v stavebníctve vzhľadom na uzavretosť údajov objavuje čoraz viac nových prístupov a technik na kontrolu a správu týchto rôznorodých údajov, hoci údaje v stavebných projektoch sú v podstate rovnaké ako v iných oblastiach. Zatiaľ čo iné odvetvia si úspešne vystačili so štandardizovanými prístupmi k spracovaniu údajov, v stavebníctve sa naďalej vyvýhajú nové a jedinečné formáty údajov, požiadavky a koncepcie validácie.

Metódy a nástroje používané na zber, prípravu a analýzu údajov v stavebníctve by sa nemali zásadne lísiť od tých, ktoré používajú odborníci v iných odvetviach hospodárstva.

V tomto odvetví sa vytvoril osobitný terminologický ekosystém, ktorý si vyžaduje kritickú reflexiu a prehodnotenie:

- Formát STEP je umiestnený pod novým názvom IFC, doplnený o kategorizáciu stavieb, bez zohľadnenia obmedzení samotného formátu STEP.
- Parametrický formát IFC sa používa v procesoch dátovej komunikácie napriek tomu, že chýba jednotné geometrické jadro potrebné na vizualizáciu a výpočty.
- Prístup k databázam CAD -systémom sa propaguje pod pojmom "BIM" bez toho, aby sa diskutovalo o vlastnostiach týchto databáz a prístupe k nim.
- Predajcovia propagujú interoperabilitu prostredníctvom formátov IFC a USD, často bez toho, aby ich uviedli do praxe, pričom používajú nákladné reverzné inžinierstvo, s ktorým sami zápasia.
- Pojmy LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD sa používajú univerzálne na opis tých istých parametrov entity bez odkazu na nástroje na modelovanie a overovanie, ktoré sa dlhodobo používajú v iných odvetviach.

Stavebníctvo dokazuje, že všetko vyššie uvedené, hoci to znie zvláštne, je v stavebníctve možné - najmä ak je hlavným cieľom speňažiť každú fázu spracovania údajov prostredníctvom predaja špecializovaných služieb a softvéru. Z obchodného hľadiska na tom nie je nič zlé. Otvorenou otázkou však zostáva, či takéto skratky a prístupy súvisiace s CAD (BIM) skutočne prinášajú pridanú hodnotu a zjednodušujú odborné procesy.

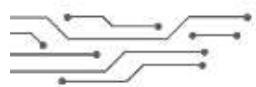
V stavebníctve takýto systém funguje, pretože samotné odvetvie dosahuje väčšinu svojich špekulatívnych ziskov práve v tomto labyrinte systémov a skratiek. Spoločnosti, ktoré majú záujem o transparentné procesy a otvorené údaje, sú zriedkavé. Táto zložitá situácia bude pravdepodobne pretrvávať donekonečna - až kým zákazníci, klienti, investori, banky a súkromný kapitál nezačnú požadovať jasnejšie a informovanejšie prístupy k správe informácií.

V odvetví sa nahromadilo nadmerné množstvo skratiek, ale všetky rôznej miere opisujú rovnaké procesy a požiadavky na údaje. Ich skutočná užitočnosť pri zjednodušovaní pracovných postupov zostáva otázna.

Zatial' čo pojmy a marketingové skratky prichádzajú a odchádzajú, samotné procesy overovania požiadaviek na údaje zostanú navždy neoddeliteľnou súčasťou obchodných procesov. Namiesto toho, aby vytváralo ďalšie a ďalšie špecializované formáty a predpisy, malo by stavebníctvo siahnuť po nástrojoch, ktoré sa už osvedčili v iných oblastiach, ako sú financie, priemysel a IT.

Množstvo pojmov, skratiek a formátov vytvára ilúziu hlboko prepracovaných procesov digitálnej výstavby. Marketingové pojmy a zložitá terminológia však často skrývajú jednoduchú, ale nepríjemnú pravdu: údaje zostávajú ľažko dostupné, nedostatočne zdokumentované a pevne viazané na konkrétné softvérové riešenia.

Aby sme sa dostali z tohto začarovaného kruhu skratiek a formátov pre formáty, je potrebné pozerať sa na systémy CAD (BIM) nie ako na magické nástroje na správu informácií, ale ako na to, čím v skutočnosti sú - špecializované databázy. A práve cez túto prizmu možno pochopiť, kde sa končí marketing a začína skutočná práca s informáciami.



## KAPITOLA 6.4.

### PARAMETRIZÁCIA NÁVRHU A POUŽITIE LLM PRE CAD PREVÁDZKU

#### Ilúzia jedinečnosti údajov CAD (BIM): cesta k analytike a otvoreným formátom

Moderné platformy CAD (BIM) výrazne zmenili prístup k riadeniu informácií o projektovaní a výstavbe. Kým predtým sa tieto nástroje používali najmä na vytváranie výkresov a 3D modelov, dnes fungujú ako plnohodnotné úložiská projektových údajov. V rámci koncepcie jednotného zdroja pravdy sa parametrický model čoraz viac stáva hlavným a často jediným zdrojom informácií o projekte, čím sa zabezpečuje jeho integrita a relevantnosť počas celého životného cyklu projektu.

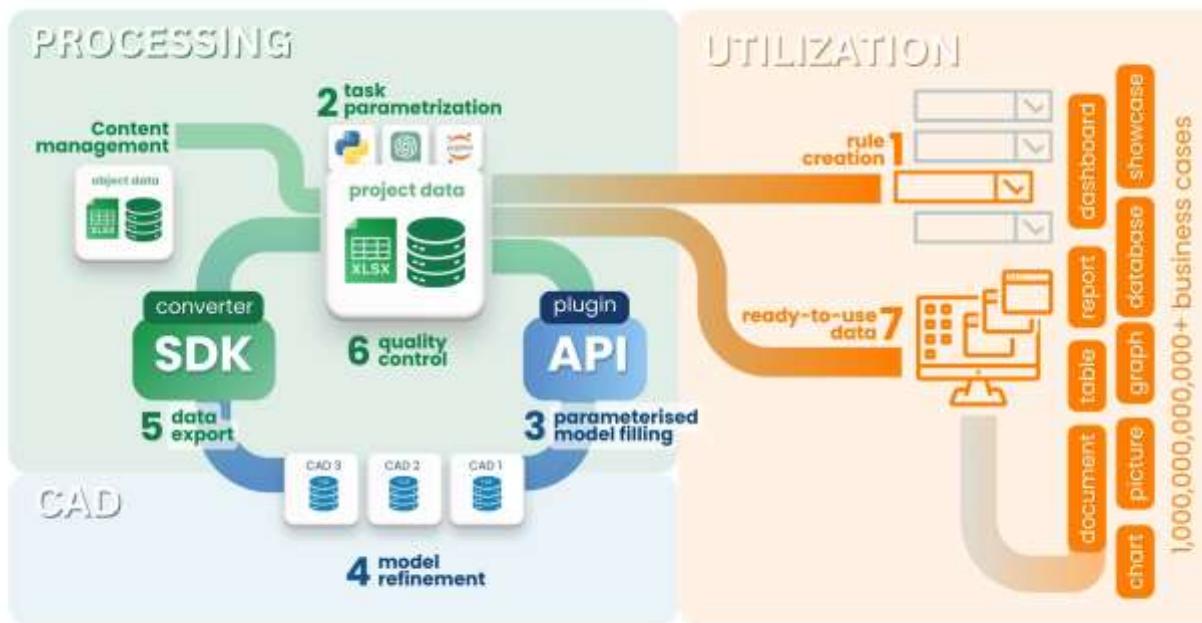
Hlavným rozdielom medzi platformami CAD - (BIM-) a inými systémami na správu stavebných údajov je potreba špecializovaných nástrojov a rozhraní API na prístup k informáciám (jediný zdroj pravdy). Tieto databázy nie sú univerzálne v tradičnom zmysle: namiesto otvorenej štruktúry a flexibilnej integrácie predstavujú uzavreté prostredie, pevne spojené s konkrétnou platformou a formátom.

Napriek zložitosti práce s údajmi CAD je dôležitejšia otázka, ktorá presahuje rámec technickej realizácie: čo sú vlastne databázy CAD (BIM)? Na zodpovedanie tejto otázky je potrebné ísiť nad rámec obvyklých skratiek a pojmov, ktoré vnučujú vývojári softvéru. Namiesto toho je vhodné zameriť sa na podstatu práce s informáciami o projekte: údaje a ich spracovanie.

Obchodný proces v stavebníctve sa nezačína prácou v nástrojoch CAD - alebo BIM -, ale tvorbou požiadaviek na projekt a modelovaním údajov. Najprv sa definujú parametre úlohy: zoznam entít, ich východiskové charakteristiky a hraničné hodnoty, ktoré je potrebné zohľadniť pri riešení konkrétnej úlohy. Až potom sa v systémoch CAD (BIM) vytvárajú modely a prvky na základe zadaných parametrov

Proces, ktorý predchádza tvorbe informácií v databázach CAD - (BIM-), je úplne rovnaký ako proces modelovania údajov, ktorý bol podrobne rozobrany v štvrtej časti knihy a v kapitole "Modelovanie údajov: konceptuálny, logický a fyzický model" (obr. 4.3-1).

Podobne ako pri dátovom modelovaní vytvárame požiadavky na údaje, ktoré chceme neskôr spracovať v databáze, pre manažérov databáz CAD vytvárame požiadavky na návrh vo forme niekoľkých stĺpcov tabuľky alebo zoznamov dvojíc kľúč-hodnota (obr. 6.4-1, kroky 1-2). A až na základe týchto počiatocných parametrov pomocou API automaticky alebo ručne projektant vytvára (alebo skôr spresňuje) objekty v CAD- (BIM) databázach (kroky 3-4), po ktorých sa opäť kontroluje ich súlad s počiatocnými požiadavkami (kroky 5-6). Tento proces - definovanie → tvorba → validácia → úprava (kroky 2-6) - sa opakuje iteratívne, kým kvalita údajov, rovnako ako pri modelovaní údajov, nedosiahne požadovanú úroveň pre cieľový systém - dokumenty, tabuľky alebo informačné panely (krok 7).

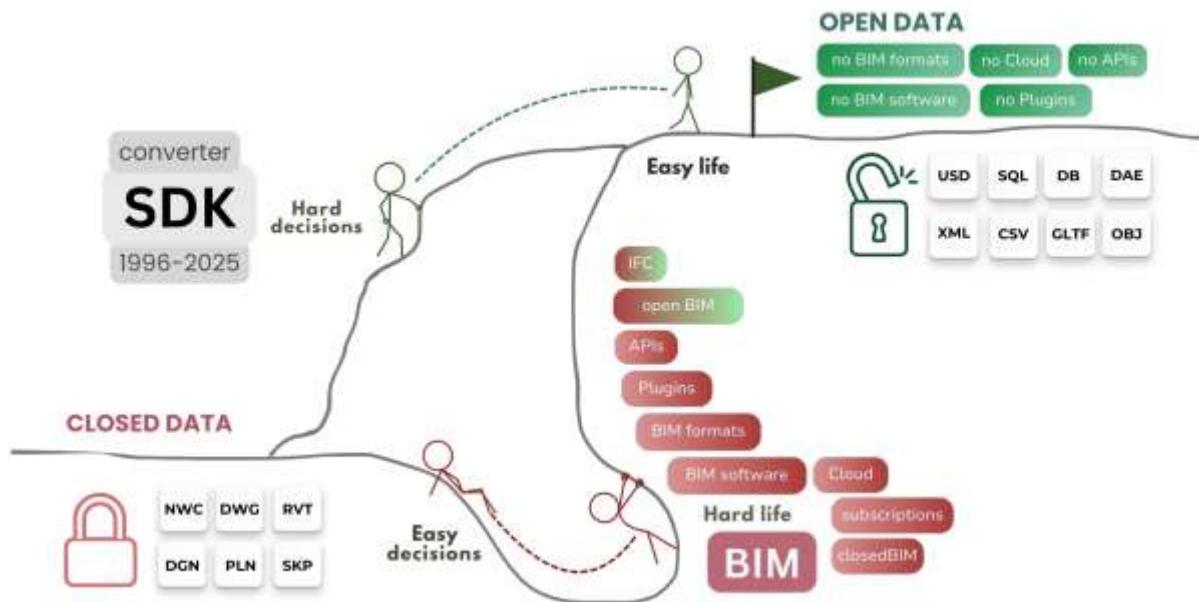


Obr. 6.4-1 Cyklus informačného nasýtenia databáz pre podnikové procesy pri realizácii stavebných projektov.

Ak považujeme CAD (BIM) za mechanizmus prenosu parametrov vo forme súboru dvojíc kľúč-hodnota generovaných na základe požiadaviek definovaných mimo návrhového prostredia (obr. 6.4-1, kroky 1-2), tăžisko diskusie sa presúva od konkrétnych softvérových riešení a ich obmedzení k základnejším aspektom - štruktúre údajov, dátovým modelom a požiadavkám na údaje. V podstate hovoríme o parametrickej saturácii databázy a klasickom procese modelovania údajov (kroky 2 - 3 a 5 - 6). Jediným rozdielom je, že vzhľadom na uzavretosť CAD-dátabáz a osobitosti používaných formátov je tento proces sprevádzaný používaním špecializovaných BIM-nástrojov. Vyvstáva otázka: v čom spočíva jedinečnosť BIM, ak v iných odvetviach neexistujú podobné prístupy?

Za posledných 20 rokov bol BIM umiestnený do pozície viac ako len jedného zdroja údajov. Balík CAD-BIM sa často predáva ako parametrický nástroj s prirodzene integrovanou databázou [64], ktorý dokáže automatizovať procesy navrhovania, modelovania a riadenia životného cyklu stavebných projektov. V skutočnosti sa však BIM stal skôr nástrojom na udržanie používateľov na platforme dodávateľa než vhodnou metódou riadenia údajov a procesov.

V dôsledku toho sú údaje CAD- (BIM-) izolované v rámci ich platform a skrývajú informácie o projekte za vlastnými API a geometrickými jadrami. To pripravilo používateľov o možnosť nezávislého prístupu k databázam a získavania, analyzovania, automatizácie a prenosu údajov do iných systémov, čím sa obchádzajú ekosystémy dodávateľov.



Obr. 6.4-2 V stavebníctve si moderné formáty vyžadujú sofistikované geometrické jadrá, každoročne aktualizované API a špeciálne licencie pre CAD -(BIM-) softvér.

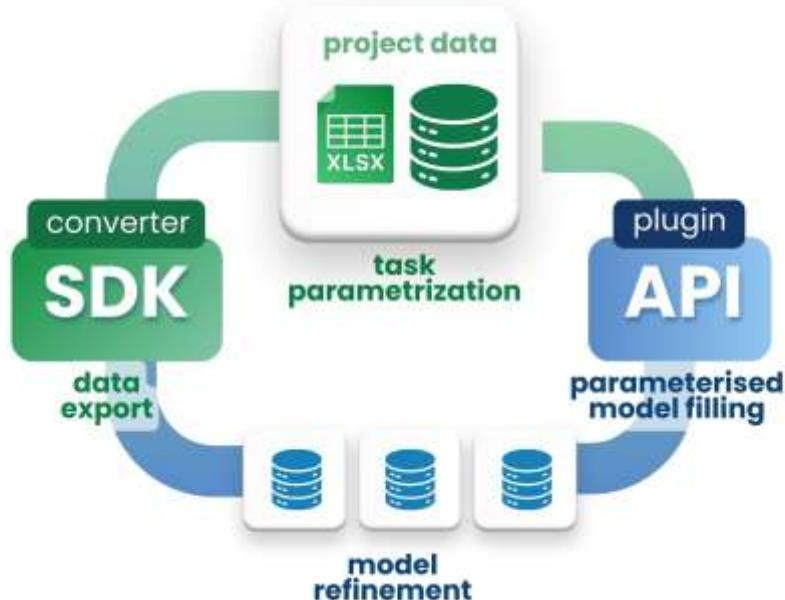
Spoločnosti, ktoré pracujú s modernými CAD nástrojmi, by mali používať rovnaký prístup k práci s údajmi, aký v praxi používajú všetci samotní výrobcovia CAD bez výnimky: transformáciu údajov pomocou SDK - nástrojov reverzného inžinierstva, proti ktorým výrobcovia CAD bojujú od roku 1995 [75]. Ak máme plný prístup k databáze CAD a používame nástroje reverzného inžinierstva, môžeme získať [127] plochú množinu entít s atribútmi a exportovať ich do ľubovoľného vhodného otvoreného formátu (obr. 6.4-2), vrátane geometrie aj parametrov konštrukčných prvkov. Tento prístup zásadne mení paradigmu práce s informáciami - od súborovo orientovanej k dátovo orientovanej architektúre:

- Dátové formáty, ako napríklad RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET a iné, obsahujú identické informácie o prvkoch tohto istého projektu. To znamená, že znalosť konkrétneho formátu a jeho schémy by nemala byť prekážkou pri práci so samotnými údajmi.
- Údaje z ľubovoľných formátov možno spojiť do jednej otvorenej štruktúrovanej a granulovanej štruktúry (obrázok 9.1-10), ktorá obsahuje trojuholníkovú geometriu MESH a vlastnosti všetkých objektových entít bez obmedzení geometrických jadier.
- Analýza údajov sa snaží o univerzálnosť: pomocou otvorených údajov môžete pracovať s údajmi o projekte bez ohľadu na použitý formát.
- Minimalizácia, ako aj závislosť od rozhraní API a doplnkov dodávateľov: práca s údajmi už nie je závislá od znalostí API.

Ked" sa požiadavky na a CAD -údaje transformujú do ľahko analyzovateľných štruktúrovaných reprezentáčnych formátov - vývojári už nie sú závislí od špecifických dátových schém a uzavretých ekosystémov.

## Návrh prostredníctvom parametrov: budúcnosť CAD a BIM

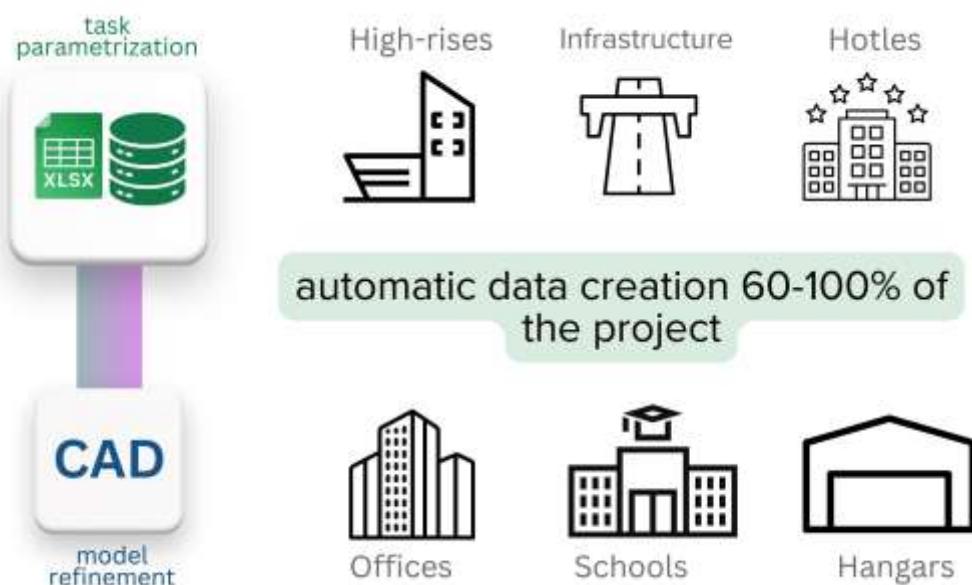
Žiadny stavebný projekt na svete sa ešte nezačal v programe CAD. Skôr ako výkres alebo model nadobudne v CAD-e svoju podobu, prechádza fázou konceptualizácie (obr. 6.4-1, fázy 1 - 2), kde sa pozornosť sústredí na parametre, ktoré definujú základnú myšlienku a logiku budúceho objektu. Táto etapa zodpovedá konceptuálnej úrovni v dátovom modelovaní (obr. 4.3-6). Parametre môžu existovať výlučne v mysli projektanta, ale v ideálnom prípade sú usporiadané vo forme štruktúrovaných zoznamov, tabuľiek alebo uložené v databázach (obr. 6.4-3), čo umožňuje transparentnosť, reproducibilnosť a ďalšiu automatizáciu procesu návrhu.



Obr. 6.4-3 Proces navrhovania je iteračný proces napĺňania databázy CAD informáciami zvonku pomocou požiadaviek v hodnotovom reťazci.

Pred začatím samotného CAD modelovania (logická a fyzická fáza dátového modelovania (obr. 4.3-7)) je dôležité definovať hraničné parametre, ktoré slúžia ako základ projektu. Tieto atribúty sa podobne ako ostatné požiadavky zhromažďujú na samom konci reťazca využívania údajov (napr. systémov) a ich prostredníctvom sa už definujú obmedzenia, ciele a kľúčové charakteristiky budúcich objektov v projekte.

Samotné modelovanie možno plne automatizovať o 60 až 100% pomocou nástrojov na parametrické modelovanie (obr. 6.4-3), ak sú požiadavky dobre definované. Len čo je projekt opísaný vo forme parametrov, jeho tvorba sa stáva technicky realizateľnou, napríklad pomocou vizuálnych programovacích jazykov, ako je Grasshopper Dynamo, zabudovaných v moderných CAD -prostrediach alebo bezplatných riešení v Blenderi, UE, Omniverse.



Obr. 6.4-4 Väčšina typizovaných projektov sa dnes už vytvára úplne automaticky vďaka parametrickým programovacím nástrojom.

Veľké priemyselné a typizované projekty už dnes nevznikajú rukami dizajnérov, ale prostredníctvom parametrických nástrojov a vizuálneho programovania. To umožňuje vytvoriť model na základe údajov, a nie na základe subjektívnych rozhodnutí konkrétneho projektanta alebo manažéra.

*Obsah predchádza dizajnu. Dizajn bez obsahu nie je dizajn, ale dekorácia [128].*

- Jeffrey Zeldman, webový dizajnér a podnikateľ

Proces sa nezačína kreslením alebo 3D - modelovaním, ale vytvorením požiadaviek. Práve požiadavky určujú, aké prvky budú v projekte použité, aké údaje je potrebné preniesť do iných oddelení a systémov. Len existencia štruktúrovaných požiadaviek umožňuje automatickú pravidelnú kontrolu modelov (napríklad aj každých 10 minút bez toho, aby sa projektant vyrušoval od svojej práce).

Možno sa v budúcnosti systém CAD- (BIM-stane len rozhraním na napĺňanie databázy a nebude záležať na tom, v akom nástroji CAD modeluje (fyzická úroveň)).

Podobne v strojárstve sa často používa 3D modelovanie, ktoré však nie je nevyhnutným alebo povinným prvkom projektu. Vo väčšine prípadov postačuje klasická 2D dokumentácia a na jej základe sa vytvára potrebný informačný model. Tento model je zostavený z komponentov štruktúrovaných podľa priemyselných noriem a obsahuje všetky potrebné informácie na pochopenie konštrukcie a organizácie výroby. Informačný model závodu sa potom použije na vytvorenie informačného modelu závodu, do ktorého sa pridajú konkrétné výrobky a vývojové diagramy, ktoré sú už orientované na potreby technológov. Celý proces je možné organizovať bez zbytočnej zložitosti, bez zaťažovania systému 3D grafikou tam, kde neprináša skutočné výhody.

Je dôležité pochopiť, že samotný 3D model a CAD -systém by nemali hrať hlavnú úlohu - je to len nástroj na kvantitatívnu a geometrickú analýzu. Všetky ostatné parametre, okrem geometrie, ktoré popisujú entitu, by mali byť uložené a spracované mimo prostredia CAD, ak je to možné (BIM).

Navrhovanie prostredníctvom parametrov nie je len trend, ale nevyhnutná budúcnosť stavebníctva. Namiesto ručného vytvárania zložitých 3D -modelov budú projektanti pracovať s údajmi, overovať ich a automatizovať procesy, čím sa stavebníctvo priblíži svetu programovania. Postupom času budú procesy projektovania postavené na princípoch vývoja softvéru:

- Vytvorenie požiadaviek → Vytvorenie modelu → Nahrávanie na server → Overenie zmeny → Pull request
- Požiadavka Pull automaticky vykoná kontrolu modelu na základe požiadaviek, ktoré boli vytvorené pred alebo počas procesu návrhu.
- Po kontrole kvality údajov a schválení sa zmeny implementujú do projektu, spoločnej databázy alebo sa automaticky prenesú do iných systémov.

Už teraz sa v strojárstve takéto konštrukčné zmeny začínajú vytvorením oznámenia o zmene. Podobný systém čaká aj stavebníctvo: projektovanie bude iteračný proces, v ktorom je každý krok podporený parametrickými požiadavkami. Takýto systém umožní projektantom vytvárať automatické kontroly a automatické žiadosti o stiahnutie špecifických požiadaviek.

Projektant budúcnosti je v prvom rade operátor s údajmi, nie manuálny modelár. Jeho úlohou je naplniť projekt parametrickými entitami, kde geometria je len jedným z atribútov.

Práve pochopenie významu modelovania údajov, klasifikácie a štandardizácie, ktoré boli podrobne rozobrané v predchádzajúcich kapitolách knihy, bude zohrávať dôležitú úlohu pri transformácii. Návrhové predpisy budúcnosti budú formalizované ako dvojice kľúč-hodnota parametrov vo forme XLSX alebo XML -schém.

Budúcnosť stavebníctva je o zhromažďovaní údajov, ich analýze, overovaní a automatizácii procesov pomocou analytických nástrojov. BIM (alebo CAD) nie je konečným cieľom, ale len etapou vývoja. Keď si odborníci uvedomia, že môžu pracovať priamo s údajmi, obchádzajúc tradičné nástroje CAD, samotný pojem "BIM" postupne ustúpi pojmom používania štruktúrovaných a granulovaných údajov o stavebných projektoch.

Jedným z kľúčových faktorov urýchľujúcich túto transformáciu bol vznik veľkých jazykových modelov (LLM) a nástrojov na nich založených. Tieto technológie menia spôsob, akým sa pracuje s návrhovými údajmi, a umožňujú prístup k informáciám bez potreby hlbokých znalostí rozhraní API alebo riešení dodávateľa. Vďaka LLM sa proces vytvárania požiadaviek a interakcie s údajmi CAD stáva intuitívnym a prístupným.

## Vznik LLM v procesoch spracovania údajov v CAD

Okrem vývoja nástrojov na prístup k databázam CAD a otvorených a zjednodušených formátov CAD sa objavujú aj nástroje LLM (Large Language Models), ktoré prinášajú revolúciu v spracovaní údajov o projektovaní. Kým v minulosti sa prístup k informáciám uskutočňoval najmä prostredníctvom zložitých rozhraní a vyžadoval si programátorské zručnosti a znalosť API, teraz je možné komunikovať s údajmi pomocou prirodzeného jazyka.

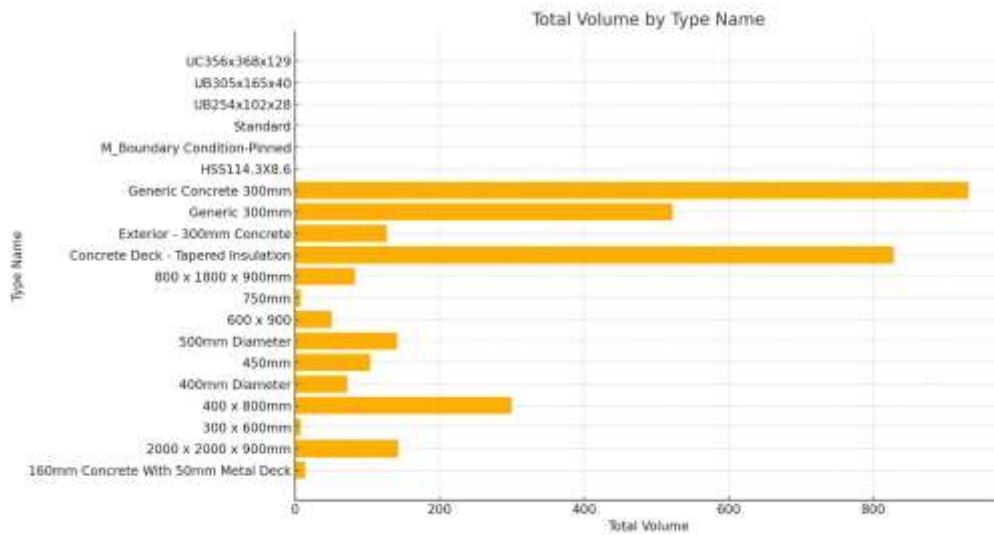
Inžinieri, manažéri a projektanti bez technického vzdelenia môžu získať potrebné informácie z projektových údajov formulovaním dotazov v bežnom jazyku. Za predpokladu, že údaje sú štruktúrované a prístupné (obr. 4.1-13), stačí položiť v LLM chate otázku typu: "Ukážte v tabuľke so zoskupením podľa typu všetky steny s objemom väčším ako 10 metrov kubických" - a model automaticky prevedie tento dotaz do jazyka SQL alebo kódu v Pandas, pričom vygeneruje súhrnnú tabuľku, graf alebo dokonca hotový dokument.

Nižšie uvádzame niekoľko reálnych príkladov interakcie LLM -modelov s návrhovými údajmi reprezentovanými v rôznych formátoch CAD- (BIM-).

- 💡 Príklad dotazu v LLM chate na projekt CAD vo formáte RVT po konverzii (obr. 4.1-13) do tabuľkového dátového rámca (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo iného):

Zoskupenie údajov v Dataframe získaných zo súboru RVT podľa "názvu typu" pri sčítaní parametra "Objem" a zobrazenie počtu prvkov v skupine. A toto všetko prosím zobrazte ako horizontálny histogram bez nulových hodnôt.

- 💡 Odozva LLM ako horizontálny stĺpcový graf (formát PNG):



Obrázok 6.4-5 Namiesto 17 kliknutí myšou alebo 40 riadkov kódu pomocou zásuvných modulov v LLM okamžite načítame tabuľku QTO prostredníctvom textového dotazu.

- 💡 Na vytvorenie tabuľky QTO typov stien s celkovou plochou a množstvom z kategórie "Steny" sformulujme textový dotaz pre LLM -chat:

Z dátového rámca projektu vyberte len tie položky, ktoré majú v parametri "Kategória" hodnotu "OST\_Steny", zoskupte ich podľa "Názvu typu", spočítajte hodnotu stĺpca "Plocha", pripočítajte množstvo a zobrazte ich v tabuľke odstránením nulových hodnôt.

- 💡 Odpoveď LLM vo forme hotovej tabuľky QTO:

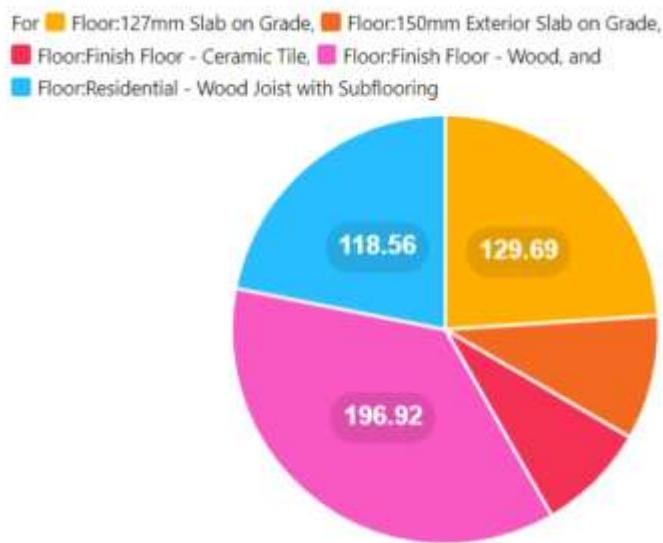
Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall_sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

Obr. 6.4-6 Vytvorenie tabuľky QTO v prirodzenom jazyku poskytuje rovnako kvalitný výsledok ako pri použití nástrojov CAD - (BIM- ).

- 💡 Požiadajme o projekt vo formáte IFC po konverzii na tabuľkový dátový rámec a zadajte podobný textový dotaz do ľubovoľného rozboru LLM:

Vezmite z projektu len tie položky, ktoré majú v parametri "Parent" (Nadriadený) hodnoty Level 1 a Level 2 (Úroveň 1 a 2), a položky, ktoré majú v parametri "Category" (Kategória) hodnoty IfcSlab, potom tieto položky zoskupte podľa parametra "ObjectType" (Typ objektu), scítajte hodnoty v parametri "PSet\_RVT \_Dimensions Area" (Oblasť rozmerov) a zobrazte ich ako koláčový graf.

💡 Odpoveď LLM ako hotový koláčový graf skupín prvkov z údajov IFC :



Obrázok 6.4-7 Výsledkom dopytu na údaje IFC v štruktúrovanom formáte môže byť akýkoľvek typ grafu, ktorý je vhodný na pochopenie údajov.

Za každým z výsledných hotových riešení (Obr. 6.4-5 - Obr. 6.4-7) sa skrýva tucet riadkov kódu jazyka Python využívajúceho knižnicu Pandas. Výsledný kód je možné skopírovať z chatu LLM a použiť v ľubovoľnom lokálnom alebo online IDE na získanie identických výsledkov mimo chatu LLM.

V tom istom chate LLM môžeme pracovať nielen s projektmi získanými z 3D CAD (BIM) formátov, ale aj s plochými výkresmi vo formáte DWG, na ktoré sa môžeme pýtať v chate LLM a zobraziť napríklad údaje o skupinách prvkov vo forme čiar alebo 3D geometrie po konverzii do štruktúrovanej podoby.

## Automatizovaná analýza súborov DWG pomocou LLM a Pandas

Proces spracovania údajov zo súborov DWG - vzhľadom na neštruktúrovanú povahu informácií - bol vždy zložitou úlohou, ktorá si vyžadovala špecializovaný softvér a často manuálnu analýzu. S rozvojom umelej inteligencie a nástrojov LLM, však bolo možné mnohé kroky tohto, dnes väčšinou manuálneho procesu, automatizovať. Uvažujme o skutočnej Potrubí požiadaviek na LLM (v tomto príklade ChatGPT) na prácu s výkresmi DWG, ktoré umožňujú pracovať s projektom:

- Filtrovanie údajov DWG podľa vrstvy, ID a súradníc
- Vizualizácia geometrie prvkov
- Automatické anotovanie výkresov na základe parametrov
- Rozšírenie polylinií stien do vodorovnej roviny
- Vytváranie interaktívnych 3D -vizualizácie rovinnych údajov

### ■ Štruktúrovanie a analýza konštrukčných údajov bez zložitých nástrojov CAD

V našom prípade sa proces budovania Pipeline začína sekvenčným generovaním kódu prostredníctvom LLM. Najprv sa vygeneruje dotaz opisujúci úlohu. ChatGPT vygeneruje kód v jazyku Python, ktorý sa vykoná a analyzuje, pričom výsledok sa zobrazí vo vnútri chatu. Ak výsledok nie je podľa očakávania, požiadavka sa opraví a proces sa opakuje

**Potrubie** je postupnosť automatizovaných krokov vykonávaných na spracovanie a analýzu údajov. V takomto procese každý krok prijíma údaje ako vstup, vykonáva transformáciu a výsledok odovzdáva ďalšiemu kroku.

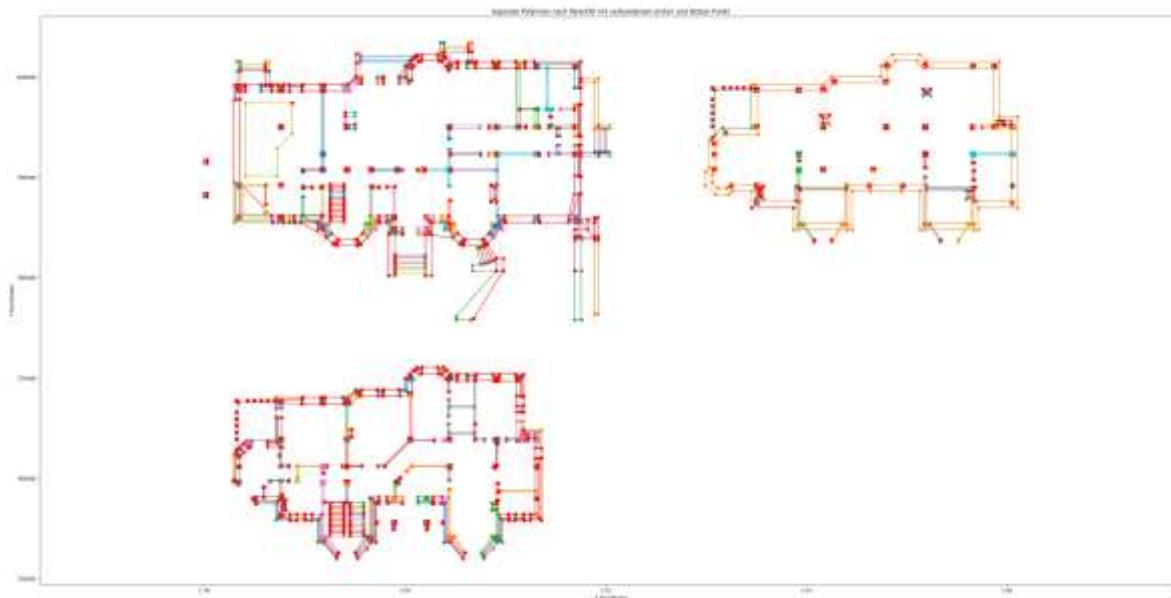
Po získaní požadovaného výsledku sa kód skopíruje z LLM a vloží sa do kódu vo forme blokov v niektorom z vhodných IDE, v našom prípade na platforme Kaggle.com. Výsledné fragmenty kódu sa spájajú do jednej pipeline, ktorá automatizuje celý proces - od načítania údajov až po ich finálnu analýzu. Tento prístup umožňuje rýchly vývoj a škálovanie analytických procesov bez hlbokých znalostí programovania. Úplný kód všetkých nižšie uvedených fragmentov spolu s ukážkovými dotazmi možno nájsť na platforme Kaggle.com vyhľadaním "DWG Analyse with ChatGPT | DataDrivenConstruction" [129]. [129].

Proces práce s DWG údajmi začnime po prevode do štruktúrovanej podoby (obr. 4.1-13), a to klasickým krokom - zoskupením a filtrovaním všetkých výkresových údajov, potrebných pre našu úlohu stenových prvkov, konkrétnie polylinií (parameter "ParentID" umožňuje zoskupiť línie do skupín), ktoré majú v parametri (stĺpci dátového rámca) "Layer" reťazcovú hodnotu obsahujúcu nasledujúcu kombináciu písmen (RegEx) - "wall".

- ▣ Ak chcete získať kód pre podobnú úlohu a výsledok vo forme obrázka, mali by ste napísať nasledujúci dotaz v LLM:

Najprv skontrolujte, či dátový rámec získaný z DWG obsahuje definované stĺpce: "Layer", "ID", "ParentID" a "Point". Potom odfiltrujte ID zo stĺpca "Layer", ktoré obsahujú reťazec "wall". Nájdite položky v stĺpco "ParentID", ktoré zodpovedajú týmto identifikátorom. Definujte funkciu na vyčistenie a rozdelenie údajov v stĺpco "Point". To zahrňa odstránenie zátvoriek a rozdelenie hodnôt na súradnice "x", "y" a "z". Vykreslite údaje pomocou matplotlib. Pre každé jedinečné "ParentID" nakreslite samostatnú polyliniu spájajúcu súradnice "Point". Ak je to možné, uistite sa, že prvý a posledný bod sú spojené. Nastavte príslušné štítky a nadpisy a zabezpečte, aby osi x a y mali rovnaké mierky.

- ▣ Odpoveď LLM vám poskytne hotový obrázok, za ktorým sa skrýva kód Pythonu, ktorý ho vygeneroval:



Obr. 6.4-8 Kód LLM extrahol všetky čiary vrstvy "stena" zo súboru DWG, vymazal ich súradnice a skonštruoval polylinie pomocou jednej z knižníc Python.

- Teraz pridajme k líniám parameter plochy, ktorý má každá polylinkia vo svojich vlastnostiach (v jednom zo stĺpcov dátového rámca):

Teraz získajte len jedno "ParentID" z každej polylinky - nájdite toto ID v stĺpci "ID", vezmiete hodnotu "Area", vydel'te ju 1 000 000 a pridajte túto hodnotu do grafu

- 💡 V odpovedi LLM sa zobrazí nový graf, v ktorom bude mať každá polylínia nadpis so svojou plochou:

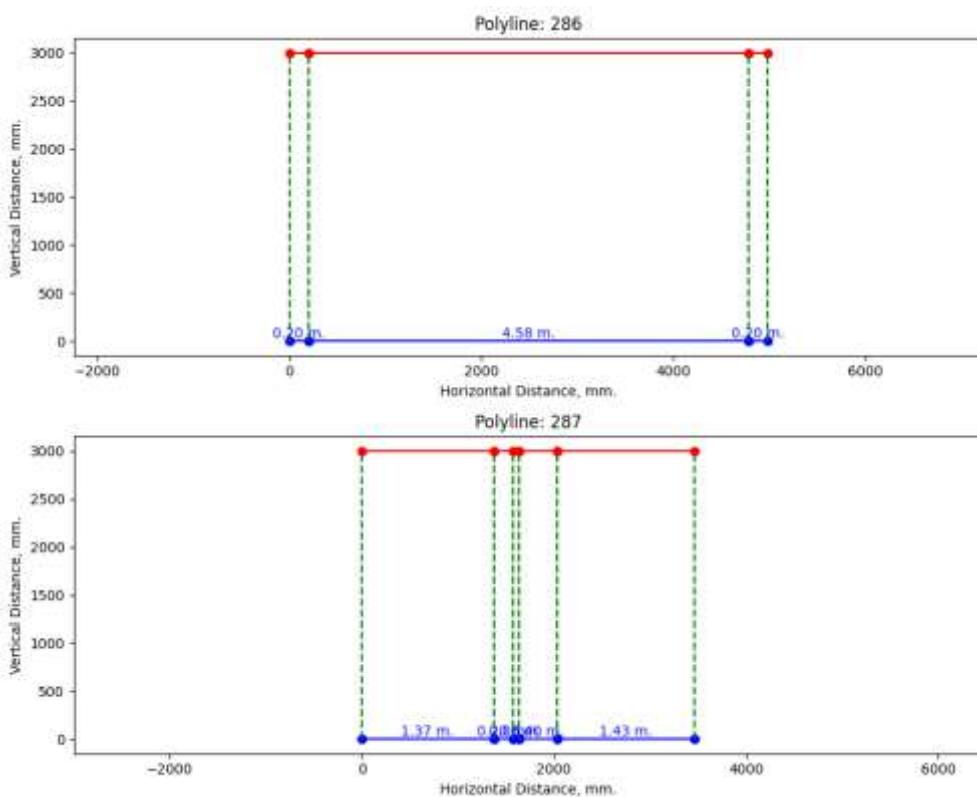


Obr. 6.4-9 LLM má pridaný kód, ktorý preberá hodnoty plochy pre každú polyliniu a pridáva ich do obrázka s vizualizáciou čiar.

- 💡 Potom prevedieme každú polyliniu na vodorovnú čiaru, pridáme rovnobežnú čiaru vo výške 3000 mm a spojíme ich do jednej roviny, aby sme týmto spôsobom zobrazili rozloženie povrchov stenových prvkov:

Musíte zobrať všetky prvky zo stĺpca "Layer" s hodnotou "wall". Vezmite tieto ID ako zoznam zo stĺpca "ID" a nájdite tieto ID z celého rámca údajov v stĺpci "ParentID". Všetky prvky sú línie, ktoré sú spojené do jednej polylinie. Každá línia má inú geometriu x, y prvého bodu v stĺpci "Point" (Bod). Musíte postupne vziať každú polyliniu a z bodu 0,0 vodorovne nakresliť dĺžku každého segmentu z polylinie. dĺžku každého segmentu polylinie do jednej čiary. Potom nakreslite presne tie isté čiary, len o 3000 vyššie, a spojte všetky body do jednej roviny.

- ❷ Odozva LLM vypíše kód, ktorý vám umožní vykresliť výkresy stien v rovine:

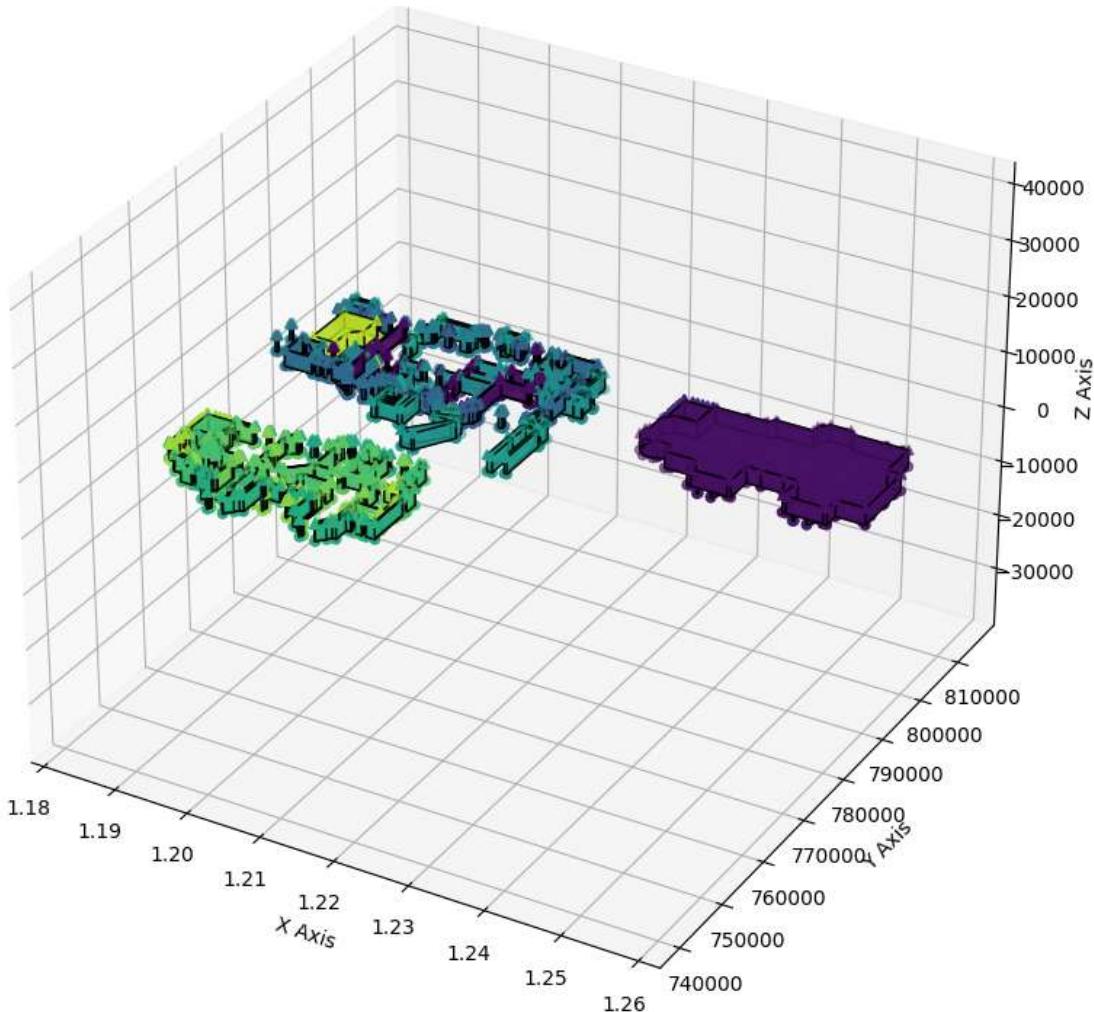


Obr. 6.4-10 Každú polylíniu pomocou podnetov premeníme na layout, ktorý vizualizuje roviny stien priamo v chate LLM.

- ❸ Teraz prejdime od 2D projekcie k 3D - modelovanie stien z plochých čiar spojením hornej a dolnej vrstvy polylinií:

Vizualizujte prvky stien v 3D, spájajúce polylnie vo výškach  $z = 0$  a  $z = 3000$  mm.  
Vytvorenie uzavretej geometrie reprezentujúcej steny budovy. Použite 3D grafický nástroj Matplotlib.

- LLM vygeneruje interaktívny 3D -graf, v ktorom bude každá polylínia reprezentovaná ako súbor rovín. Používateľ sa bude môcť voľne pohybovať medzi prvkami pomocou počítačovej myši a skúmať model v 3D režime kopírovaním kódu z chatu do IDE:



Obr. 6.4-11 LLM pomohol vytvoriť kód [129] na vizualizáciu plochých kresliacich čiar do 3D zobrazenia, ktoré možno preskúmať v 3D prehliadači v rámci IDE.

Na vytvorenie logického a reprodukovateľného potrubia - od počiatočnej konverzie a načítania súboru DWG - až po konečný výsledok - sa odporúča po každom kroku skopírovať vytvorený blok kódu LLM - do IDE. Takto si výsledok nielen skontrolujete v chate, ale ho aj okamžite spustíte vo svojom vývojovom prostredí. To vám umožní postupné zostavovanie procesu, jeho ladenie a prispôsobovanie podľa potreby.

Kompletný kód Pipeline všetkých fragmentov (obrázky 6.4-8 až obrázky 6.4-11) spolu so vzorovými dotazmi nájdete na platforme Kaggle.com vyhľadaním "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. Na stránke Kaggle si môžete nielen prezrieť kód a použité výzvy, ale

aj skopírovať a otestovať celý Pipeline s pôvodnými DWG dátovými rámcami v cloude zadarmo bez toho, aby ste museli inštalovať akýkoľvek ďalší softvér alebo samotné IDE.

Prístup uvedený v tejto kapitole umožňuje plne automatizovať kontrolu, spracovanie a generovanie dokumentov na základe DWG -projektov. Vyvinutý Pipeline je vhodný tak na spracovanie jednotlivých výkresov, ako aj na dávkové spracovanie desiatok, stoviek a tisícok DWG-súborov s automatickým generovaním potrebných zostáv a vizualizácií pre každý projekt.

Proces možno organizovať sekvenčne a transparentne: najprv sa údaje zo súboru CAD automaticky prevedú do formátu XLSX, potom sa načítajú do dátového rámca, nasleduje zoskupovanie, kontrola a generovanie výsledkov - všetko sa realizuje v jednom zápisníku Jupyter alebo skripte Python v ľubovoľnom populárnom prostredí IDE. V prípade potreby možno tento proces ľahko rozšíriť prostredníctvom integrácie so systémami na správu projektovej dokumentácie: súbory CAD možno automaticky načítať podľa zadaných kritérií, výsledky možno vrátiť späť do systému na ukladanie údajov a používateľia môžu byť upozornení, keď sú výsledky pripravené - prostredníctvom e-mailu alebo messengerov.

Používanie chatov a agentov LLM na prácu s návrhovými údajmi znižuje závislosť od špecializovaných programov CAD a umožňuje vykonávať analýzu a vizualizáciu architektonických návrhov bez nutnosti manuálnej interakcie s rozhraním - bez klikania myšou a zapamätania si zložitej navigácie v menu.

Každým dňom bude stavebníctvo čoraz viac počuť o LLM, granulárne štruktúrovaných údajoch, DataFrame a stĺpcových databázach. Jednotné dvojrozumné DataFrame vytvorené z rôznych databáz a formátov CAD, budú ideálnym palivom pre moderné analytické nástroje, s ktorými aktívne pracujú špecialisti v iných odvetviach.

Samotný proces automatizácie sa výrazne zjednoduší - namiesto študovania API uzavretých niche produktov a písania zložitých skriptov na analýzu alebo transformáciu parametrov teraz bude stačiť formulovať úlohu vo forme súboru jednotlivých textových príkazov, ktoré sa zložia do požadovaného Pipeline alebo Workflow-procesu pre požadovaný programovací jazyk, ktorý beží zadarmo na takmer akomkoľvek zariadení. Už žiadne čakanie na nové produkty, formáty, pluginy alebo aktualizácie od výrobcov nástrojov CAD- (BIM-). Inžinieri a stavbári budú môcť samostatne pracovať s údajmi pomocou jednoduchých, bezplatných a ľahko pochopiteľných nástrojov, pričom im budú pomáhať chaty a agenti LLM.

## **Ďalšie kroky: prechod od uzavretých formátov k otvoreným údajom**

Pri práci s návrhovými údajmi budúcnosti je nepravdepodobné, že by niekto skutočne potreboval rozumieť geometrickým jadrám proprietárnych nástrojov alebo sa učiť stovky nekompatibilných formátov obsahujúcich rovnaké informácie. Bez pochopenia toho, prečo je prechod na otvorené štruktúrované údaje dôležitý, je však ľahké argumentovať v prospech používania nových bezplatných nástrojov, otvorených údajov a prístupov, ktoré pravdepodobne nebudú presadzovať výrobcovia softvéru.

V tejto kapitole sme sa venovali kľúčovým vlastnostiam údajov CAD (BIM), ich obmedzeniam a možnostiam a tomu, že napriek marketingovým sľubom predajcov sa inžinieri a projektanti každý deň stretávajú s ťažkosťami pri získavaní, prenose a analýze informácií o návrhu. Pochopenie architektúry týchto systémov a oboznámenie sa s alternatívnymi prístupmi - založenými na otvorených formátoch a automatizácii prostredníctvom LLM - môže výrazne uľahčiť život aj jednotlivým odborníkom, nehovoriac o firmách. Na zhrnutie tejto časti stojí za to zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať prístupy rozoberané v tejto časti na vaše každodenné úlohy:

- Rozšírite svoj súbor nástrojov na prácu s údajmi o projektoch
  - Preskúmajte dostupné zásuvné moduly a nástroje na získavanie údajov z používaných systémov CAD - (BIM-)
  - Preskúmajte dostupné SDK a API, ktoré vám umožnia automatizovať extrakciu údajov z uzavretých formátov bez nutnosti manuálneho otvárania špecializovaného softvéru.
  - Ovládanie základných zručností pri práci s otvorenými neparametrickými geometrickými formátmi (OBJ, glTF, USD, DAE) a zodpovedajúcimi otvorenými knižnicami
  - Skúste vymysliť systém na ukladanie metadát projektu oddelene od geometrie mimo riešení CAD (BIM), aby sa zjednodušila analýza a integrácia s inými systémami.
  - Používajte LLM na automatizáciu konverzie údajov medzi formátmami
- Vytvorenie vlastných procesov na spracovanie informácií o projekte
  - Začať opisovať úlohy a požiadavky na modelovanie prostredníctvom parametrov a ich hodnôt v jednoduchých a štruktúrovaných formátoch
  - Vytvorenie osobnej knižnice skriptov alebo blokov kódu pre často vykonávané operácie
- Podporujte používanie otvorených štandardov vo svojej práci
  - Vyzvať kolegov a partnerov, aby zdieľali údaje v otvorených formátoch, ktoré nie sú obmedzené ekosystémom dodávateľov softvéru.
  - Demonštrovať výhody používania štruktúrovaných údajov na konkrétnych príkladoch
  - Iniciovať diskusie o problémoch s uzavretými formátmami a možných riešeniach.

Aj keď nemôžete zmeniť politiku vašej spoločnosti týkajúcu sa platformou CAD - (BIM-), osobné pochopenie zásad práce s projektovými údajmi v otvorených formátoch vám umožní výrazne zvýšiť efektivitu vašej práce. Vytvorením vlastných nástrojov a metód na extrakciu a transformáciu údajov z rôznych formátov nielenže optimalizujete svoje pracovné postupy, ale získate aj flexibilitu, ktorá vám umožní obísť obmedzenia štandardných softvérových riešení.



## VII ČASŤ

# ROZHODOVANIE ZALOŽENÉ NA ÚDAJOCH, ANALYTIKA, AUTOMATIZÁCIA A STROJOVÉ UČENIE

Siedma časť je zameraná na analýzu údajov a automatizáciu procesov v stavebníctve. Rozoberá, ako sa údaje stávajú základom pre rozhodovanie, a vysvetľuje zásady vizualizácie informácií pre efektívnu analýzu. Podrobne sú opísané kľúčové ukazovatele výkonnosti (KPI), metódy hodnotenia návratnosti investícií (ROI) a vytváranie informačných panelov na monitorovanie projektov. Osobitná pozornosť sa venuje procesom ETL (Extract, Transform, Load) a ich automatizácii pomocou potrubí (Pipeline) na premenu rôznorodých údajov na štruktúrované informácie na analýzu. Diskutuje sa o nástrojoch na orchestráciu pracovných tokov, ako sú Apache Airflow, Apache NiFi a n8n, ktoré umožňujú vytvárať automatizované dátové potrubia bez hlbokých znalostí programovania. Významnú úlohu zohrávajú veľké jazykové modely (LLM) a ich využitie na zjednodušenie analýzy údajov a automatizáciu rutinných úloh.

## KAPITOLA 7.1.

### **ANALÝZA ÚDAJOV A ROZHODOVANIE ZALOŽENÉ NA ÚDAJOCH**

Po krokoch zberu, štruktúrovania, čistenia a overovania informácií vznikol ucelený a analyzovateľný súbor údajov. Predchádzajúce časti knihy sa zaobrali systematizáciou a štruktúrovaním heterogénnych zdrojov - od dokumentov PDF a textových záznamov zo stretnutí až po modely CAD a geometrické údaje. Podrobne je opísaný proces kontroly a zosúladenia informácií s požiadavkami rôznych systémov a klasifikátorov, odstránenie duplicit a nekonzistentností.

Všetky výpočty vykonávané na týchto údajoch (tretia, štvrtá časť knihy) - od jednoduchých transformácií až po výpočty ukazovateľov času, nákladov a ESG (piata časť) - sú agregovanými analytickými úlohami. Tvoria základ pre pochopenie aktuálneho stavu projektu, posúdenie jeho parametrov a následné rozhodovanie. V dôsledku toho sa údaje v dôsledku výpočtov menia zo súboru rôznorodých záznamov na zvládnuteľný zdroj schopný odpovedať na kľúčové obchodné otázky.

V predchádzajúcich kapitolách boli podrobne opísané procesy zberu údajov a kontroly kvality pre použitie v typických obchodných prípadoch a procesoch špecifických pre stavebnictvo. Analytika v tomto kontexte je v mnohých ohľadoch podobná aplikáciám v iných odvetviach, ale má niekoľko špecifických vlastností.

V nasledujúcich kapitolách bude podrobne opísaný proces analýzy údajov vrátane krokov automatizácie od prvotného získania informácií a ich transformácie až po následný prenos do cieľových systémov a dokumentov. Najprv bude predstavená teoretická časť, ktorá sa zameria na vybrané aspekty rozhodovania založeného na údajoch. Potom bude v nasledujúcich kapitolách nasledovať praktická časť týkajúca sa automatizácie a konštrukcie ETL -Pipeline.

#### **Údaje ako zdroj pri rozhodovaní**

Rozhodovanie založené na údajoch je často opakujúci sa proces a začína systematickým zhromažďovaním informácií z rôznych informačných zdrojov. Podobne ako v prírodnom cykle jednotlivé dátové prvky a celé informačné systémy postupne padajú do pôdy - hromadia sa v informačných úložiskách spoločnosti (obrázok 1.3-2). Časom sa tieto údaje, podobne ako opadané lístie a konáre, premenia na hodnotný materiál. Mycélium dátových inžinierov a analytikov organizuje a pripravuje informácie na budúce použitie a mení opadané údaje a systémy na cenný kompost, aby z nich vyrástli nové výhonky a nové systémy (obr. 1.2-5).

Trendy v širokom využívaní analytiky v rôznych odvetviach znamenajú začiatok novej éry, v ktorej sa práca s údajmi stáva základom profesionálnej činnosti (obr. 7.1-1). Je dôležité, aby sa odborníci v stavebnictve prispôsobili týmto zmenám a boli pripravení na prechod do novej éry - éry dát a analytiky

Ručné presúvanie údajov medzi tabuľkami a manuálne vykonávanie výpočtov sa postupne stáva minulosťou a ustupuje automatizácii, analýze toku údajov, analytike a strojovému učeniu. Tieto nástroje sa stávajú kľúčovými prvkami moderných systémov na podporu rozhodovania.

V knihe spoločnosti McKinsey "Reštartovanie. sprievodca prekonávaním konkurencie vo veku digitálnych technológií a umelej inteligencie" [130] sa uvádza prieskum, ktorý sa uskutočnil v roku 2022 na vzorke 1 330 vrcholových manažérov z rôznych regiónov, odvetví a funkčných oblastí [130], cituje prieskum uskutočnený v roku 2022 s 1 330 vrcholovými manažérmi z rôznych regiónov, odvetví a funkčných oblastí. Podľa jeho výsledkov 70% vedúcich pracovníkov využíva pokročilú analytiku na generovanie vlastných nápadov a 50% zavádzajú umelú inteligenciu na zlepšenie a automatizáciu rozhodovacích procesov.



Obr. 7.1-1 Analýza údajov a analytika je hlavným nástrojom na zvýšenie rýchlosťi rozhodovania v podniku.

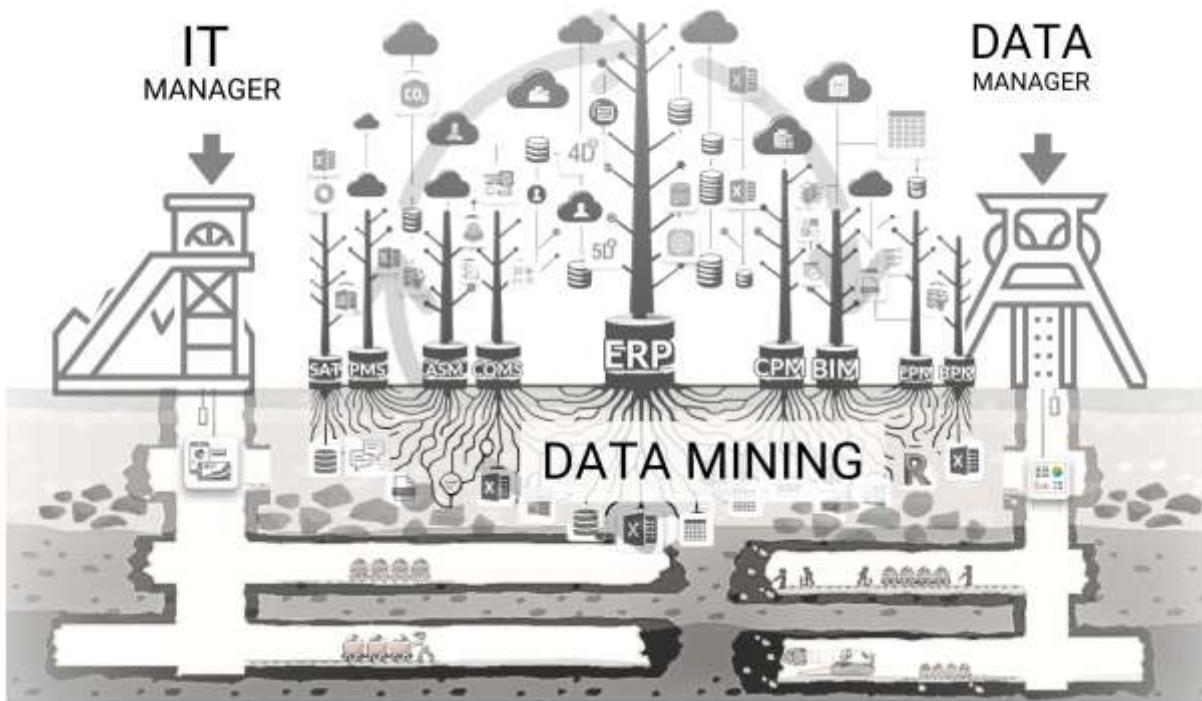
Analýza údajov, podobne ako šírenie mycélia, preniká do humusu minulých rozhodnutí, pomáha spájať jednotlivé systémy a vedie manažérov k cenným poznatkom. Tieto poznatky, podobne ako živiny z rozložitých stromov dátových systémov, napájajú nové rozhodnutia v podniku, vedú k efektívnym zmenám a rastu kvality informácií, podobne ako nové výhonky a klíčky vyrastajúce z bohatej a zdravej pôdy (obrázok 1.2-5).

*Čísla majú dôležitý príbeh. Spoliehajú sa na to, že im dáte jasný a presvedčivý hlas [131].*

- Stephen Few, odborník na vizualizáciu údajov

V stredne veľkých a malých podnikoch je dnes práca na získavaní a príprave informácií na ďalšiu analýzu mimoriadne prácy proces (obrázok 7.1-2), porovnatelný s ťažbou uhlia v osiemnásťom storočí. Donedávna bola práca na získavaní a príprave údajov vyhradená skôr pre dobrodruhov pracujúcich vo vysoko špecializovanej medzere s malým a obmedzeným súborom nástrojov na prácu s rôznymi typmi údajov z neštruktúrovaných, voľne štruktúrovaných, zmiešaných a uzavretých zdrojov.

Rozhodovacie orgány a manažéri často nemajú skúsenosti s heterogénnymi údajmi a systémami, ale musia na ich základe prijímať rozhodnutia. V dôsledku toho sa rozhodovanie založené na údajoch v modernom stavebnictve v posledných desaťročiach menej podobá na automatizovaný proces a viac na niekoľkodňovú manuálnu prácu baníka v prvých uhoľných baniach.



Obr. 7.1-2 V procese dolovania údajov prechádzajú experti zložitou cestou prípravy údajov - od čistenia až po štruktúrovanie na následnú analýzu.

Hoci sú moderné metódy získavania údajov v stavebnictve určite pokročilejšie ako primitívne techniky baníkov z 12. storočia, stále ide o zložitú a vysoko rizikovú úlohu, ktorá si vyžaduje značné zdroje a odborné znalosti, ktoré si môžu dovoliť len veľké spoločnosti. Procesy získavania a analýzy údajov z nahromadeného dedičstva minulých projektov donedávna vykonávali prevažne veľké, technologicky vyspelé spoločnosti, ktoré údaje sústavne zhromažďovali a uchovávali desiatky rokov.

Predtým zohrávali vedúcu úlohu v oblasti analytiky technologicky vyspelé spoločnosti, ktoré zhromažďovali údaje už desaťročia. Dnes sa situácia mení: prístup k údajom a nástrojom na ich spracovanie sa demokratizuje - predtým zložité riešenia sú teraz dostupné pre každého zadarmo.

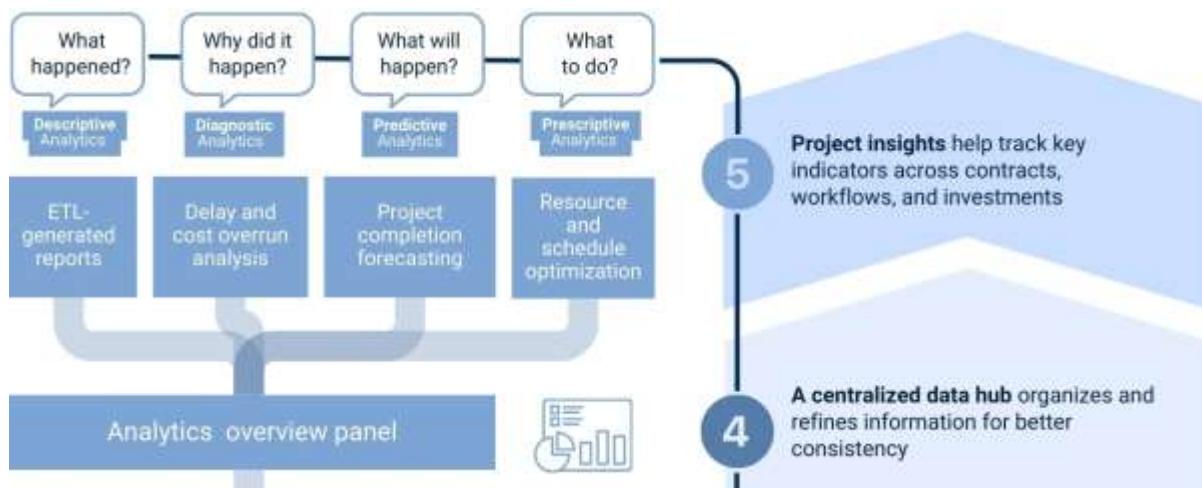
Aplikácia analytiky umožňuje spoločnostiam prijímať presnejšie a informovanejšie rozhodnutia v reálnom čase. Nasledujúca prípadová štúdia ilustruje, ako môžu historické údaje pomôcť prijímať finančne správne rozhodnutia:

- ☛ **Projektový manažér** - "Teraz je priemerná cena betónu v meste 82€ /m<sup>3</sup>, v odhade máme 95 €/m<sup>3</sup>."
- ☛ **Odhadca** - "Pri predchádzajúcich projektoch bolo prekročenie približne 15%, takže som sa stiahol."
- ☛ **Manažér údajov alebo inžinier kontroly na strane zákazníka** - "Pozrime sa na analytiku posledných troch tendrov."

Po analýze DataFrame z minulých projektov dostaneme:

- **Priemerná skutočná nákupná cena:** 84,80 /m<sup>3</sup>€
- **Priemerný pomer prekročenia výdavkov:** +4,7%
- ▲ **Odporučaná sadzba v odhade:** ~ 85 /m<sup>3</sup>€

Takéto rozhodnutie sa už nebude zakladať na subjektívnych pocitoch, ale na konkrétnych historických štatistikách, čo pomáha znižovať riziká a zvyšovať platnosť ponuky vo verejnej súťaži. Analýza údajov z minulých projektov sa stáva akýmsi "organickým hnojivom", z ktorého klíčia nové, presnejšie riešenia.



**Obrázok 7.1-3 Analýza údajov odpovedá na tri klúčové otázky: čo sa stalo, prečo sa to stalo a čo by sa malo urobiť ďalej.**

Rozhodovacie orgány a manažéri často čelia potrebe pracovať s heterogénnymi údajmi a systémami bez dostatočných technických znalostí. V takýchto situáciach je vizualizácia, jeden z prvých a najdôležitejších krovov v analytickom procese, klúčovou pomôckou pri pochopení údajov. Umožňuje prezentovať informácie vo vizuálnej a zrozumiteľnej forme.

## Vizualizácia údajov: klúč k pochopeniu a rozhodovaniu

V dnešnom stavebníctve, kde sa projektové údaje vyznačujú komplexnosťou a viacúrovňovou štruktúrou, zohráva vizualizácia klúčovú úlohu. Vizualizácia údajov umožňuje projektovým manažérom a inžinierom vizualizovať zložité vzory a trendy skryté vo veľkých, heterogénnych objemoch údajov.

Vizualizácia údajov na stránke uľahčuje pochopenie stavu projektu: pridelovanie zdrojov, vývoj nákladov alebo spotreba materiálu. Grafy a diagramy sprístupňujú a pochopiteľne zobrazujú zložité a suché informácie, čo vám umožňuje rýchlo identifikovať klúčové oblasti, ktorým je potrebné venovať pozornosť, a odhaliť potenciálne problémy.

Vizualizácia údajov nielenže uľahčuje interpretáciu informácií, ale je kľúčovým krokom v analytickom procese a informovanom rozhodovaní v oblasti riadenia, pretože pomáha odpovedať na otázky "čo sa stalo?" a "ako sa to stalo?" (obrázok 2.2-5). (Obrázok 2.2-5).

*Grafika je vizuálny nástroj na riešenie logických problémov [132].*

- Jacques Bertin, "Grafika a grafické spracovanie informácií

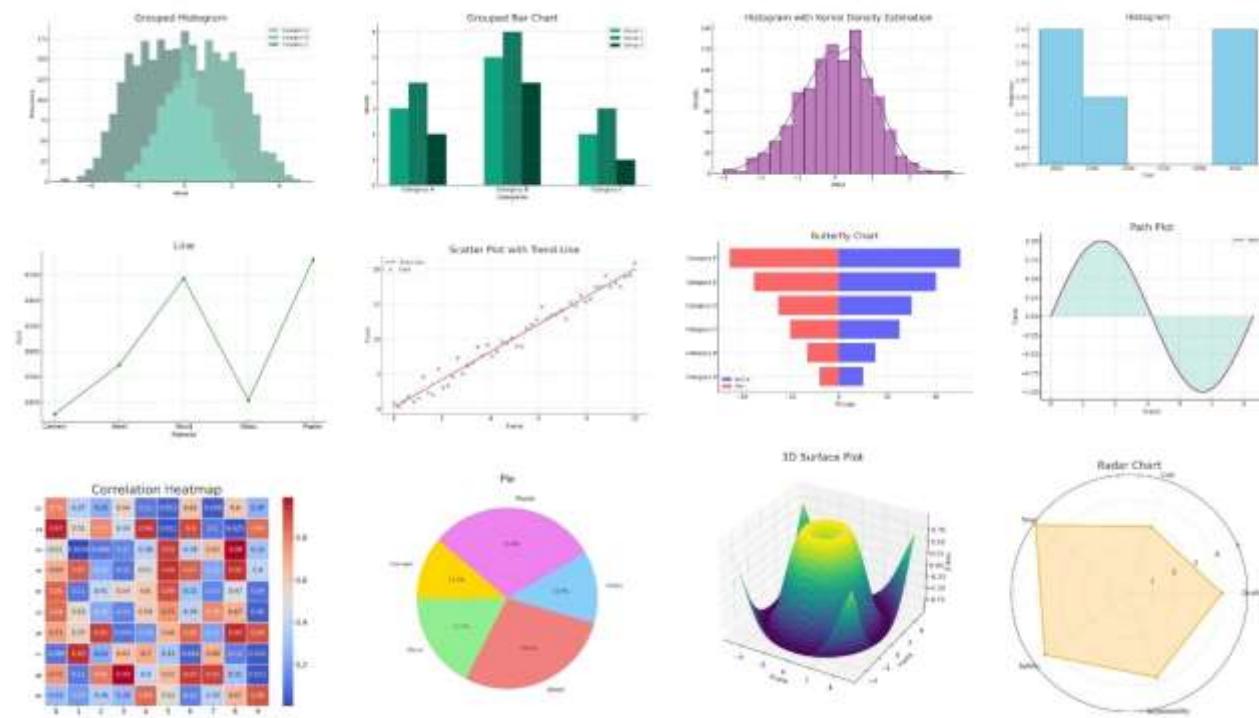
Projektoví manažéri pred priatím kľúčových rozhodnutí častejšie používajú vizuálne znázornenie údajov než suché a ťažko interpretovateľné údaje z tabuľiek alebo textových správ.

Údaje bez vizualizácie sú ako stavebný materiál chaoticky rozhádzaný po stavenisku: ich potenciál je nejasný. Až keď sa jasne vizualizujú, ako dom z tehál a betónu, je ich hodnota jasná. Kým sa dom nepostaví, nemožno povedať, či sa z hromady materiálov stane malá chatka, luxusná vila alebo mrakodrap.

Spoločnosti majú k dispozícii údaje z rôznych systémov (obrázok 1.2-4 až obrázok 2.1-10), finančné transakcie a rozsiahle textové údaje. Využitie týchto údajov na obchodné výhody je však často náročné. V takýchto situáciách sa vizualizácia stáva dôležitým nástrojom na sprostredkovanie významu údajov, ktorý pomáha prezentovať informácie vo formátoch zrozumiteľných každému odborníkovi, ako sú informačné panely, grafy a tabuľky.

Štúdia spoločnosti PwC "Čo potrebujú študenti, aby uspeli v rýchlo sa meniacom svete podnikania" (2015) zdôrazňuje [9], že úspešné spoločnosti idú nad rámec analýzy údajov a aktívne využívajú interaktívne vizualizačné nástroje, ako sú grafy, infografiky a analytické panely na podporu rozhodovania. Podľa správy - vizualizácia údajov pomáha zákazníkom pochopiť príbeh, ktorý údaje rozprávajú prostredníctvom grafov, diagramov, tabuľiek a interaktívnych dátových modelov.

Proces prevodu informácií do vizuálnych grafických foriem, ako sú tabuľky, grafy a diagramy, zlepšuje pochopenie a interpretáciu údajov ľudským mozgom (obrázok 7.1-4). To umožňuje projektovým manažérom a analytikom rýchlejšie posúdiť zložité scenáre a prijímať informované rozhodnutia na základe vizuálne rozpoznateľných trendov a vzorov namiesto intuície.



**Obrázok 7.1-4 Rôzne typy vizualizácie sú navrhnuté tak, aby pomohli ľudskému mozgu lepšie pochopiť a dať zmysel suchým informáciám čísel.**

Tvorba vizualizácií z údajov a používaniu rôznych bezplatných vizualizačných knižníc sa budeme podrobnejšie venovať v nasledujúcej kapitole o procesoch ETL.

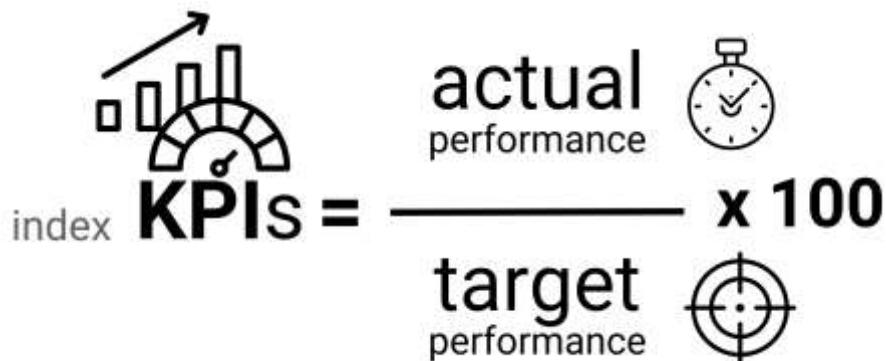
Vizualizácia sa stáva neoddeliteľnou súčasťou práce s údajmi v stavebníctve - pomáha nielen "vidieť" údaje, ale aj pochopiť ich význam v kontexte úloh riadenia. Aby však vizualizácia bola skutočne užitočná, je potrebné vopred určiť, čo presne by sa malo vizualizovať a ktoré ukazovatele sú skutočne dôležité pre hodnotenie výkonnosti projektu. Práve tu prichádzajú na rad metriky výkonnosti, ako sú KPI a ROI. Bez nich hrozí, že aj tie najkrajšie informačné panely budú len "informačným šumom".

## KPI a ROI

V súčasnom stavebníctve zohráva riadenie ukazovateľov výkonnosti (KPI a ROI) a ich vizualizácia prostredníctvom správ a informačných panelov kľúčovú úlohu pri zvyšovaní produktivity a efektívnosti riadenia projektov.

Ako v každom podnikaní, aj v stavebníctve je potrebné jasne definovať ukazovatele, podľa ktorých sa meria úspech, návratnosť investícií a výkonnosť. Pri získavaní údajov o rôznych procesoch sa organizácia, ktorá sa riadi údajmi, musí najprv naučiť identifikovať **kľúčové KPI (Key Performance Indicators)** - kvantitatívne meradlá, ktoré odrážajú mieru dosahovania strategických a prevádzkových cieľov.

Na výpočet KPI sa zvyčajne používa vzorec (obr. 7.1-5), ktorý zahŕňa skutočné a plánované ukazovatele. Napríklad na výpočet individuálneho KPI pre projekt, zamestnanca alebo proces vydelte skutočný výkon plánovaným výkonom a výsledok vynásobte 100%.



Obrázok 7.1-5 KPI sa používajú na meranie úspešnosti projektu alebo procesu pri dosahovaní klúčových cieľov.

Podrobnejšie metriky KPI možno použiť na úrovni staveniska:

- **Načasovanie klúčových miľníkov** (založenie, inštalácia, dokončenie) - umožňuje kontrolovať dodržiavanie pracovných plánov.
- **Percento prekročenia materiálu** - pomáha riadiť obstarávanie a minimalizovať plytvanie.
- **Počet neplánovaných odstávok strojov** - ovplyvňuje produktivitu a náklady.

Výber nesprávnych metrík môže viesť k chybným rozhodnutiam "čo robiť?" (obrázok 2.2-5). Ak sa napríklad spoločnosť zameria len na náklady na meter štvorcový, ale nezohľadní náklady na úpravy, úspory na materiáloch môžu viesť k nižšej kvalite a vyšším nákladom v budúcich projektoch.

Pri stanovovaní cieľov je dôležité, aby bolo jasné, čo sa meria. Nejasné formulácie vedú k nesprávnym záverom a komplikujú kontrolu. Pozrime sa na príklady úspešných a neúspešných KPI v stavebnictve.

#### Dobré klúčové ukazovatele výkonnosti:

- ❑ "Do konca roka znížte podiel rekonštrukčných prác o 10%."
- ❑ "Zvýšenie rýchlosťi montáže fasád o 15 percent bez zníženia kvality do ďalšieho štvrtroka"
- ❑ "Zníženie prestojov strojov o 20% optimalizáciou pracovných plánov do konca roka"

Tieto ukazovatele sú jasne merateľné, majú konkrétné hodnoty a časové rámce.

#### Zlé KPI:

- ❑ "Budeme stavať rýchlejšie" (O koľko rýchlejšie? Čo znamená "rýchlejšie"?).
- ❑ "Zlepšíme kvalitu konkrétnej práce" (Ako presne sa meria kvalita?)
- ❑ "Zlepšíme interakciu s dodávateľmi na stavbe" (Aké kritériá preukážu zlepšenie?)

Dobrý KPI je taký, ktorý sa dá merať a objektívne vyhodnotiť. V stavebnictve je to obzvlášť dôležité, pretože bez jasných ukazovateľov nie je možné sledovať výkonnosť a dosahovať stabilné výsledky.

Okrem KPI existuje ďalší ukazovateľ na hodnotenie efektívnosti investícií: **ROI (Return on Investment)** - ukazovateľ návratnosti investícií, ktorý vyjadruje pomer medzi ziskom a investovanými prostriedkami. ROI pomáha posúdiť, či je zavedenie nových metód, technológií alebo nástrojov opodstatnené: od digitálnych riešení a automatizácie (napr. obrázok 7.3-2) až po používanie nových stavebných materiálov. Tento ukazovateľ pomáha prijímať informované rozhodnutia o ďalších investíciách na základe ich reálneho vplyvu na ziskovosť podniku

V kontexte riadenia stavebných projektov možno ROI (návratnosť investícií) použiť ako jeden z klúčových ukazovateľov výkonnosti (KPI), ak je cieľom spoločnosti merať návratnosť investícií do projektu, technológie alebo zlepšenia procesu. Ak sa napríklad zavádzajú nová technika riadenia výstavby, ROI môže ukázať, o koľko sa vďaka nej zvýšila ziskovosť.

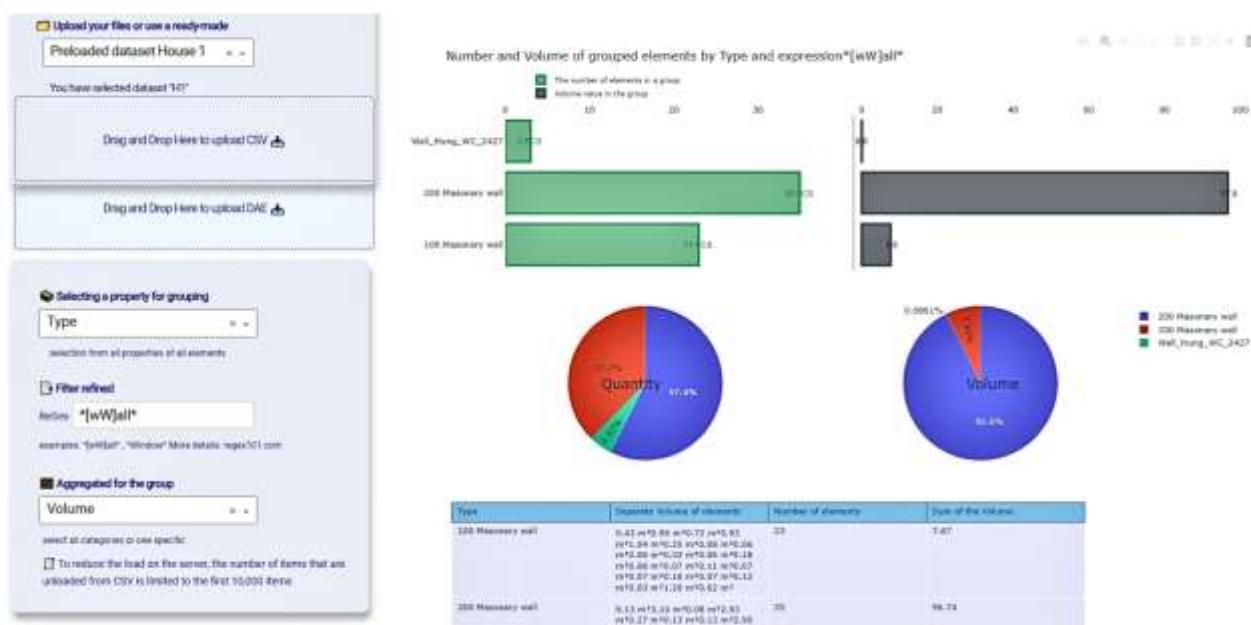
Pravidelné meranie klúčových ukazovateľov výkonnosti a návratnosti investícií na základe údajov získaných z rôznych zdrojov, ako je spotreba materiálu, počet hodín práce a náklady, umožňuje vedeniu projektu efektívne riadiť zdroje a prijímať rýchle rozhodnutia. Dlhodobé ukladanie týchto údajov umožňuje analyzovať budúce trendy a optimalizovať procesy.

Na vizualizáciu klúčových ukazovateľov výkonnosti, návratnosti investícií a ďalších ukazovateľov sa používajú rôzne grafy a tabuľky, ktoré sa zvyčajne spájajú do informačných panelov.

## **Prístrojové panely a informačné tabuľky: vizualizácia metrík pre efektívne riadenie**

Na vizualizáciu ukazovateľov a metrík sa používajú rôzne grafy a tabuľky, ktoré sa zvyčajne spájajú do prehľadov údajov a informačných panelov. Tieto informačné panely poskytujú centralizovaný pohľad na stav projektu alebo jeho častí a zobrazujú klúčové ukazovatele (ideálne v reálnom čase). Aktuálne a priebežne aktualizované dashboardy umožňujú tímu rýchlo reagovať na zmeny.

**Informačné panely** sú nástroje, ktoré vizualizujú kvantitatívne hodnotenia, aby boli ľahko dostupné a zrozumiteľné pre všetkých účastníkov projektu.



Obr. 7.1-6 Správa kľúčových ukazovateľov výkonnosti a ich vizualizácia prostredníctvom informačných panelov je kľúčom k zvýšeniu produktivity a efektívnosti projektu.

Tu je niekoľko príkladov populárnych nástrojov, v ktorých môžete vytvárať informačné panely:

- **Power BI** je nástroj od spoločnosti Microsoft na vytváranie interaktívnych zostáv a ovládacích panelov.
- **Tableau a Google Data Studio** sú výkonné nástroje na vizualizáciu údajov a vytváranie ovládacích panelov bez nutnosti písania kódu.
- **Plotly** (obrázok 7.1-6, obrázok 7.2-12) je knižnica na vytváranie interaktívnych grafov a Dash je rámcový webový aplikácií na analýzu údajov. Môžu sa použiť v kombinácii na vytvorenie interaktívnych informačných panelov.
- **Množstvo knižníč pre Python** (obrázok 7.2-9 - obrázok 7.2-11) - Python má množstvo open source a bezplatných knižníč na vizualizáciu údajov, ako napríklad Matplotlib, Seaborn, Plotly, Bokeh a ďalšie. Tie sa dajú použiť na vytváranie grafov a ich integráciu do webovej aplikácie pomocou rámcov, ako je Flask alebo Django.
- **Knižnice JavaScriptu**: umožňuje vytvárať interaktívne informačné panely pomocou knižníč JavaScriptu s otvoreným zdrojovým kódom, ako je D3.js alebo Chart.js, a integrovať ich do webovej aplikácie.

Aby ste mohli vyhodnocovať kľúčové ukazovatele výkonnosti a vytvárať informačné panely, potrebujete aktuálne údaje a jasný harmonogram zberu a analýzy informácií.

Vo všeobecnosti tvoria KPI, ROI a informačné panely v stavebnictve základ analytického prístupu k riadeniu projektov. Pomáhajú nielen monitorovať a vyhodnocovať aktuálny stav, ale poskytujú aj cenné

poznatky pre budúce plánovacie a optimalizačné procesy - procesy, ktoré priamo závisia od interpretácie údajov a kladenia správnych a včasných otázok.

## Analýza údajov a umenie klášť otázky

Interpretácia údajov je poslednou fázou analýzy, v ktorej informácie dávajú zmysel a začínajú "hovoriť". Tu sa formulujú odpovede na kľúčové otázky: "Čo robiť?" a "Ako robiť?" (obrázok 2.2-5). Táto fáza umožňuje zhrnúť výsledky, identifikovať vzory, stanoviť príčinné súvislosti a vyvodíť závery na základe vizualizácie a štatistickej analýzy.

*Možno nie je ďaleko čas, keď si uvedomíme, že na to, aby sme sa stali plnohodnotným občanom jedného z veľkých nových komplexných svetových štátov, ktoré sa teraz rozvíjajú, je rovnako potrebné vedieť počítať, myslieť v pojoch priemerov, maxím a miním, ako je dnes potrebné vedieť čítať a písat' [133].*

- Samuel S. Wilkes, citovaný v prezidentskom prejave pred Americkou štatistickou asociáciou v roku 1951

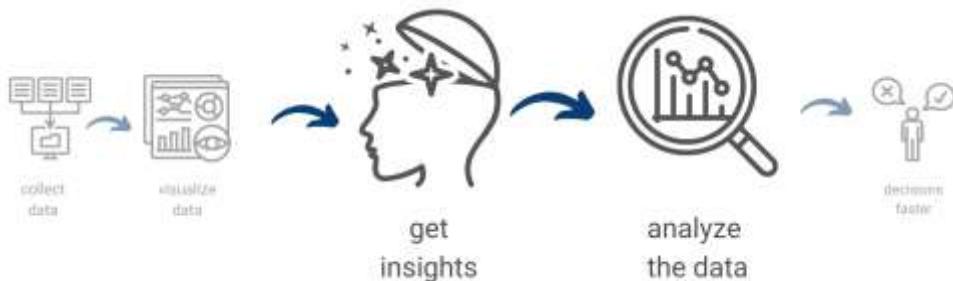
Podľa správy "Analýza údajov a umelá inteligencia pri realizácii vládnych projektov" (2024), ktorú zverejnila vláda Spojeného kráľovstva [83], môže implementácia analytiky údajov a umelej inteligencie (AI) výrazne zlepšiť procesy riadenia projektov, zvýšiť presnosť predpovedania času a nákladov, ako aj znížiť riziko a neistotu. V dokumente sa zdôrazňuje, že verejné organizácie, ktoré využívajú pokročilé analyticke nástroje, dosahujú vyššiu výkonnosť v infraštrukturálnych iniciatívach.

Moderné stavebnictvo pôsobiace vo vysoko konkurenčnom prostredí s nízkymi maržami štvrtnej priemyselnej revolúcie možno prirovnáť k vojenským operáciám. Tu prežitie a úspech podniku závisí od rýchlosť získavania zdrojov a kvalitných informácií - a teda od včasného a informovaného rozhodovania (obr. 7.1-7).

Ak je vizualizácia údajov "inteligenciou", ktorá poskytuje prehľad, potom je analýza údajov "muníciou" potrebnou na konanie. Odpovedá na otázky: "Čo robiť?" a "Ako to robiť?" a tvorí základ pre získanie konkurenčnej výhody na trhu.

Analýza mení rôznorodé údaje na štruktúrované a zmysluplné informácie, na základe ktorých sa prijímajú rozhodnutia.

Úlohou analytikov a manažérov nie je len interpretovať informácie, ale ponúkať informované rozhodnutia, identifikovať trendy, určovať vzťahy medzi rôznymi typmi údajov a kategorizovať ich v súlade s cieľmi a špecifikami projektu. Pomocou vizualizačných nástrojov a metód štatistickej analýzy menia údaje na strategický prínos pre spoločnosť.



Obr. 7.1-7 Práve analýza údajov v konečnom dôsledku mení získané informácie na zdroj pre rozhodovanie.

Na to, aby sme v procese analýzy mohli prijímať skutočne informované rozhodnutia, je potrebné naučiť sa správne formulovať otázky, ktoré sa na údaje kladú. Kvalita týchto otázok priamo ovplyvňuje hĺbku získaných poznatkov a následne aj kvalitu manažérskych rozhodnutí.

*Minulosť existuje len do tej miery, do akej je prítomná v záznamoch dneška. A to, čo tieto záznamy predstavujú, je dané otázkami, ktoré si kladieme. Neexistuje iná história ako táto [134].*

- John Archibald Wheeler, fyzik 1982

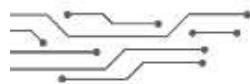
Umenie klásť hlboké otázky a kriticky myslieť je základnou zručnosťou pri práci s údajmi. Väčšina ľudí má tendenciu klásť jednoduché, povrchné otázky, ktorých zodpovedanie si nevyžaduje veľké úsilie. Skutočná analýza sa však začína zmysluplnými a premyslenými otázkami, ktoré dokážu odhaliť skryté vzťahy a príčinno-následkové súvislosti v informáciách, ktoré môžu byť skryté za viacerými vrstvami zdôvodnenia.

Podľa štúdie "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017) [135] si úspešná digitálna transformácia vyžaduje investície do analytických kapacít, programov riadenia zmien a zosúladenie obchodných cieľov s IT iniciatívami. Spoločnosti, ktoré vytvárajú kultúru založenú na údajoch, by mali investovať do kapacít dátovej analytiky a spustiť programy riadenia zmien s cieľom zaviesť nové myslenie, správanie a spôsoby práce.

Bez investícií do rozvoja analytickej kultúry, zlepšenia dátových nástrojov a školenia odborníkov budú spoločnosti nadálej riskovať, že budú prijímať rozhodnutia na základe zastaraných alebo neúplných informácií - alebo sa budú spoliehať na subjektívne názory manažérov HiPPO (obrázok 2.1-9).

Uvedomenie si dôležitosti a potreby neustále aktualizovať analytické a informačné panely nevyhnutne viedie manažment k pochopeniu dôležitosti automatizácie analytických procesov. Automatizácia zvyšuje rýchlosť rozhodovania, znížuje vplyv ľudského faktora a zabezpečuje relevantnosť údajov. S exponenciálnym rastom objemu informácií sa rýchlosť stáva nielen konkurenčnou výhodou, ale aj kľúčovým faktorom udržateľného úspechu.

Automatizácia procesov analýzy a spracovania údajov vo všeobecnosti je neoddeliteľne spojená s témou ETL (Extract, Transform, Load). Tak ako v procese automatizácie musíme údaje transformovať, v procese ETL sa údaje získavajú z rôznych zdrojov, transformujú sa podľa potrebných požiadaviek a načítavajú sa do cieľových systémov na ďalšie použitie.



## KAPITOLA 7.2.

### TOK ÚDAJOV BEZ MANUÁLNEJ NÁMAHY: PREČO JE POTREBNÝ ETL

#### Automatizácia ETL: zníženie nákladov a zrýchlenie spracovania údajov

Ked' kľúčové ukazovatele výkonnosti (KPI) prestanú rásť naprieck nárastu objemu údajov a veľkosti tímu, vedenie spoločnosti si nevyhnutne uvedomí potrebu automatizácie procesov. Skôr či neskôr sa toto uvedomenie stane podnetom na spustenie komplexnej automatizácie, ktorej hlavným cieľom je znížiť zložitosť procesov, urýchliť spracovanie a znížiť závislosť od ľudského faktora.

Podľa štúdie spoločnosti McKinsey "Ako vybudovať dátovú architektúru na podporu inovácií - dnes a zajtra" (2022) [136] získavajú spoločnosti využívajúce architektúru prúdových dát významnú výhodu, pretože môžu analyzovať informácie v reálnom čase. Streamingové technológie umožňujú priamu analýzu správ v reálnom čase a aplikáciu prediktívnej údržby vo výrobe prostredníctvom analýzy údajov zo senzorov v reálnom čase.

Zjednodušenie procesov je automatizácia, pri ktorej sa tradičné manuálne funkcie nahradzajú algoritmami a systémami.

Otázka automatizácie, alebo skôr "minimalizácie úlohy človeka pri spracovaní údajov", je pre každú spoločnosť nezvratným a veľmi citlivým procesom. Špecialisti v akejkoľvek odbornej oblasti často váhajú, či majú kolegom optimalizátorom úplne prezraditi svoje metódy a jemnosti práce, pretože si uvedomujú riziko straty zamestnania v rýchlo sa vyvíjajúcim technologickom prostredí.

*Ak si chcete získať nepriateľov, skúste veci zmeniť [137].*

- Woodrow Wilson, prejav na kongrese obchodníkov, Detroit, 1916

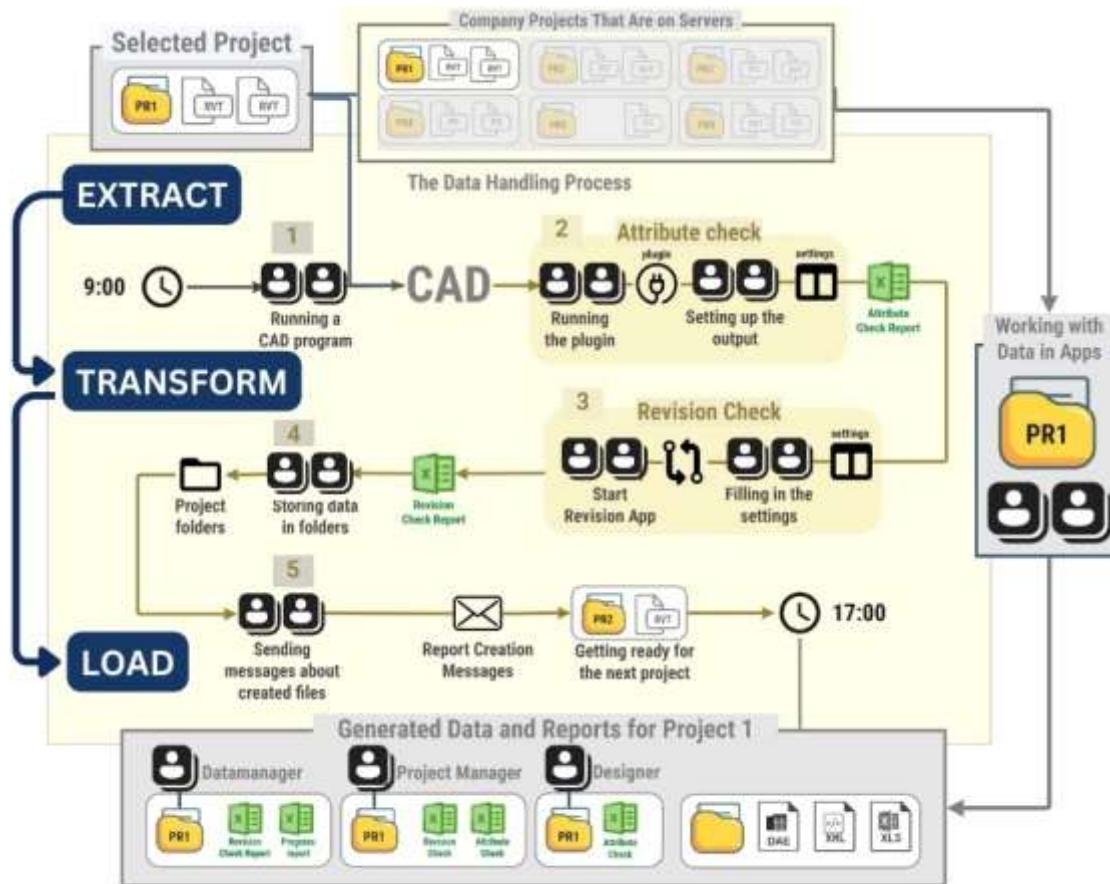
Napriek zjavným výhodám automatizácie má mnoho spoločností vo svojej každodennej praxi stále vysoký podiel manuálnej práce, najmä v oblasti technických údajov. Na ilustráciu súčasnej situácie sa pozrieme na typický príklad sekvenčného spracovania údajov v rámci takýchto procesov.

Manuálne spracovanie údajov možno ilustrovať na príklade interakcie s informáciami získanými z databáz CAD. Tradičné spracovanie údajov ("ručné" ETL -proces) v oddeleniach CAD (BIM) na vytvorenie atribútových tabuľiek alebo vytvorenie dokumentácie na základe údajov o návrhu prebieha v nasledujúcim poradí (obr. 7.2-1):

1. Manuálne **extrahovanie (Extract)**: používateľ manuálne otvorí projekt - spustením aplikácie CAD (BIM) (obr. 7.2-1 krok 1).
2. **Overenie**: Ďalší krok zvyčajne zahŕňa manuálne spustenie niekoľkých zásuvných modulov alebo pomocných aplikácií na prípravu údajov a posúdenie ich kvality (obr. 7.2-1 krok 2-3).
3. Manuálna **transformácia (Transformácia)**: po príprave sa začína spracovanie údajov, ktoré si

vyžaduje manuálnu obsluhu rôznych softvérových nástrojov, v ktorých sa údaje pripravujú na odoslanie (obr. 7.2-1 krok 4).

- Manuálne **nahrávanie (Load)**: manuálne nahrávanie konvertovaných údajov do externých systémov, dátových formátov a dokumentov (obr. 7.2-1 krok 5).



Obr. 7.2-1 Tradičné manuálne spracovanie ETL je obmedzené želанияmi a fyzickými možnosťami jednotlivých technikov.

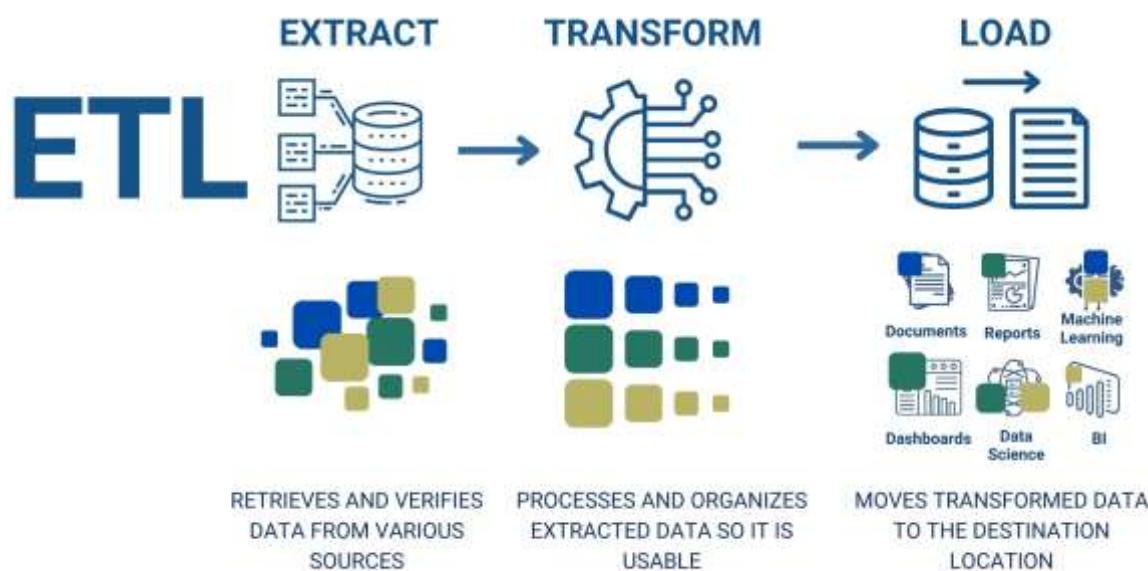
Takýto pracovný postup je príkladom klasického procesu ETL - extrakcie, transformácie a načítania (ETL). Na rozdiel od iných priemyselných odvetví, kde sú automatické potrubia ETL už dlho štandardom, v stavebníctve stále prevláda manuálna práca, ktorá spomaľuje procesy a zvyšuje náklady

**ETL (Extract, Transform, Load)** je proces extrakcie údajov z rôznych zdrojov, ich transformácie do požadovaného formátu a načítania do cieľového systému na ďalšiu analýzu a použitie.

**ETL** je proces, ktorý označuje tri kľúčové zložky spracovania údajov: Extrakcia, Transformácia a Načítanie (obrázok 7.2-2):

- **Extrahovať** - extrahovať údaje z rôznych zdrojov (súbory, databázy, API).
- **Transformácia** - čistenie údajov, agregácia, normalizácia a logické spracovanie.
- **Načítanie** - načítanie štruktúrovaných informácií do dátového skladu, zostavy alebo systému BI.

Skôr v knihe sa koncepcia ETL vyskytla len sporadicky: pri konverzii neštruktúrovaného naskenovaného dokumentu do štruktúrovanej tabuľkovej podoby (obr. 4.1-1), v súvislosti s formalizáciou požiadaviek na systematizáciu vnímania životných aj obchodných procesov (obr. 4.4-20) a pri automatizácii validácie a spracovania údajov z CAD riešení. Pozrime sa teraz na ETL podrobnejšie v kontexte typických pracovných postupov.



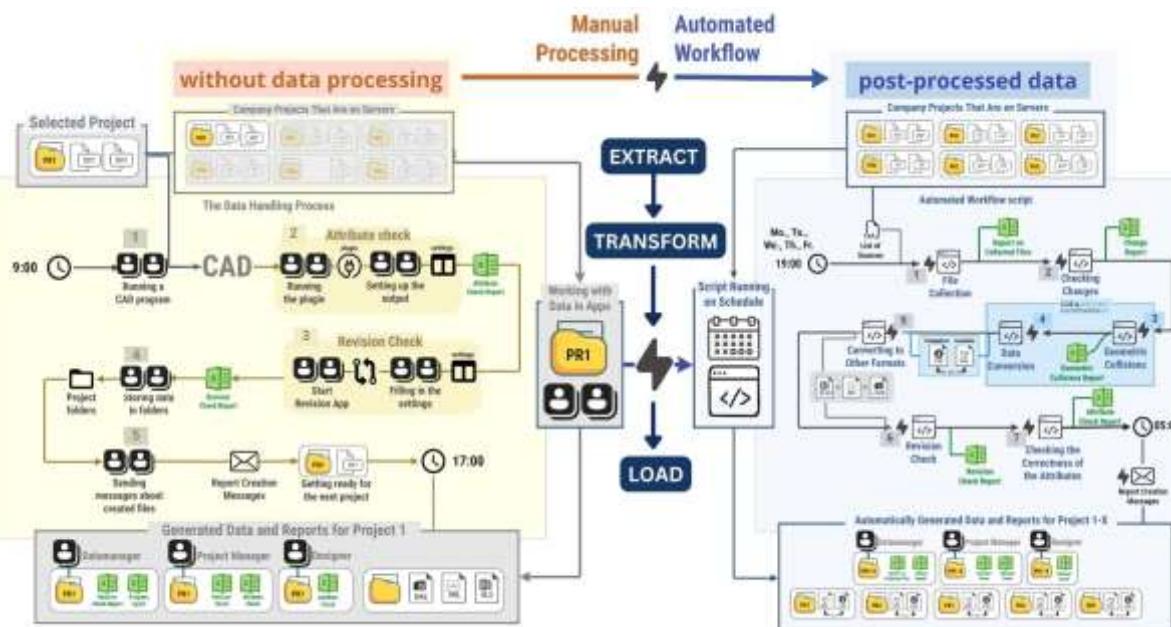
Obrázok 7.2-2 ETL automatizuje opakujúce sa úlohy spracovania údajov.

Manuálne alebo poloautomatizované ETL -proces predpokladá manažéra alebo technika, ktorý riadi všetky kroky manuálne - od zberu údajov až po generovanie správ. Takýto proces si vyžaduje značný čas, najmä ak je pracovný čas obmedzený (napr. od 9:00 do 17:00).

Spoločnosti sa často snažia vyriešiť problém nízkej efektívnosti a pomalej rýchlosťi nákupom modulárnych integrovaných riešení (ERP, PMIS, CPM, CAFM, atď.), ktoré potom ďalej rozvíjajú externí dodávateľa a konzultanti. Títo dodávateľa a externí vývojári sa však často stávajú kritickým miestom závislosti: ich technické obmedzenia priamo ovplyvňujú výkonnosť celého systému a podniku ako celku, ako je podrobne uvedené v predchádzajúcich kapitolách o proprietárnych systémoch a formátoch. Problémy, ktoré vznikajú v dôsledku fragmentácie a závislosti, boli podrobne rozobrané v kapitole "Ako sa stavebné podniky topia v dátovom chaoze".

Ak spoločnosť nie je pripravená implementovať veľkú modulárnu platformu od niektorého z dodávateľov, začne hľadať alternatívne spôsoby automatizácie. Jednou z nich je vývoj vlastných modulárnych otvorených ETL -konvertorov, kde sú jednotlivé fázy (extrakcia, transformácia, validácia, načítanie) implementované ako skripty vykonávané podľa plánu.

V automatizovanej verzii tohto istého pracovného postupu ETL (obr. 7.2-1) vyzerá pracovný postup ako modulárny kód, ktorý začína spracovaním údajov a ich prevodom do otvorenej štruktúrovanej podoby. Po prijatí štruktúrovaných údajov sa automaticky, podľa plánu, spúšťajú rôzne skripty alebo moduly na kontrolu zmien, transformáciu a odosielanie správ (obr. 7.2-3).



Obr. 7.2-3 Vľavo je ručné obrábanie, vpravo je automatický proces, ktorý na rozdiel od tradičného ručného obrábania nie je obmedzený možnosťami používateľa.

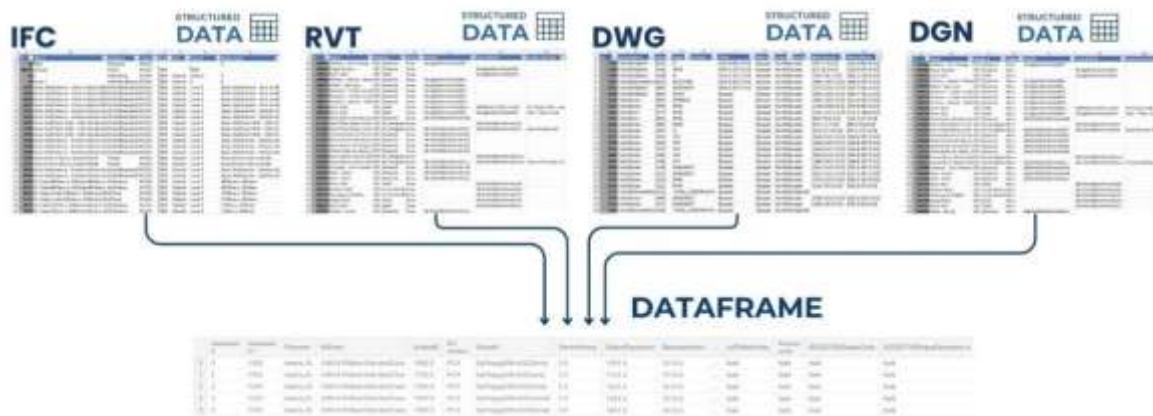
V automatizovanom pracovnom postupe sa spracovanie údajov zjednodušuje predbežným spracovaním údajov ET(L): štruktúrovaním a zjednotením.

Pri tradičných metódach spracovania pracujú odborníci s údajmi "tak, ako sú" - ako sú získané zo systémov alebo softvéru. V automatizovaných procesoch naopak údaje často najprv prechádzajú cez ETL -payplane, kde sa uvedú do konzistentnej štruktúry a formátu vhodného na ďalšie použitie a analýzu.

Uveďme si praktický príklad ETL, ktorý demonštruje proces validácie tabuľky údajov opísaný v kapitole "Validácia údajov a výsledky validácie" (obrázok 4.4-13). Na tento účel použijeme knižnicu Pandas v spojení s LLM na automatizované procesy analýzy a spracovania údajov.

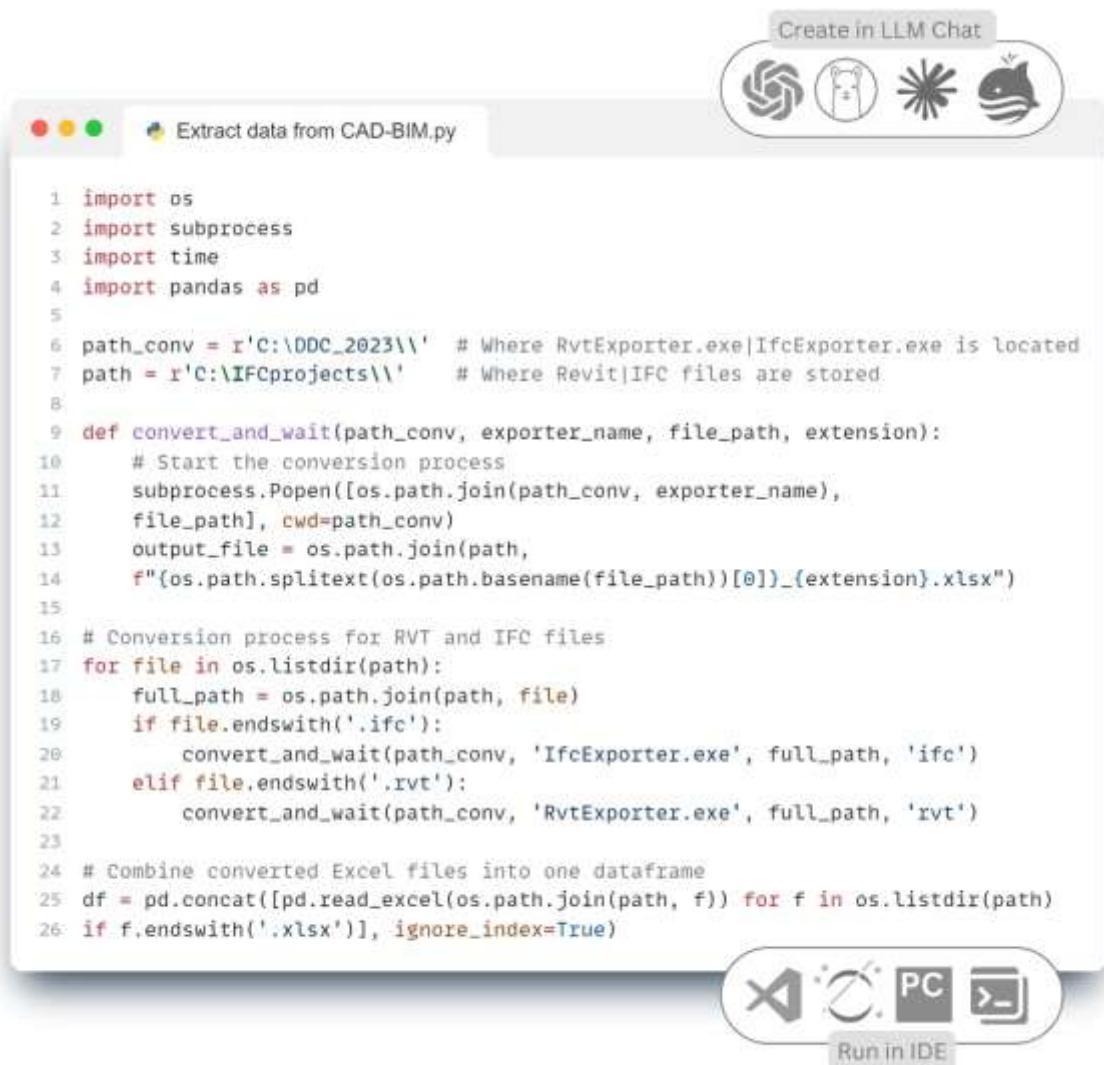
## ETL Výpis: zber údajov

Prvá fáza procesu ETL - Extract) - sa začína písaním kódu na zhromažďovanie súborov údajov, ktoré sa majú ďalej kontrolovať a spracovať. Na tento účel budeme skenovať všetky priečinky produkčného servera, zhromažďovať dokumenty určitého formátu a obsahu a potom ich previesť do štruktúrovanej podoby. Tento proces je podrobne opísaný v kapitolách "Prevod neštruktúrovaných a textových údajov do štruktúrovanej podoby" a "Prevod údajov CAD (BIM) do štruktúrovanej podoby" (obr. 4.1-1 - obr. 4.1-12).



Obr. 7.2-4 Konvertovať údaje CAD (BIM) do jedného veľkého dátového rámca, ktorý bude obsahovať všetky časti projektu.

Ako ilustračný príklad použijeme krok načítania údajov Extract a získame tabuľku všetkých projektov CAD- (BIM-) (obr. 7.2-4) používa konvertory s podporou reverzného inžinierstva [138] formáty RVT a IFC na získanie štruktúrovaných tabuľiek zo všetkých projektov a ich spojenie do jednej veľkej tabuľky DataFrame.



```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe|IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                     file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Obr. 7.2-5 Konverzia pomocou kódu Python a nástroja SDK reverzného inžinierstva pre súbory RVT a IFC do jedného veľkého štruktúrovaného (df) DataFrame.

Pandas DataFrame dokáže načítať údaje z rôznych zdrojov vrátane textových súborov CSV, tabuľiek Excel, súborov JSON - a XML -, formátov na ukladanie veľkých objemov údajov, ako sú Parquet a HDF5, a z databáz MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle a iných. Okrem toho Pandas podporuje načítavanie údajov z rozhraní API, webových stránok, clouдовých služieb a systémov ukladania, ako sú Google BigQuery, Amazon Redshift a Snowflake.

- ❷ Ak chcete napísať kód na pripojenie a zber informácií z databáz, pošlite podobnú textovú požiadavku na chat LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akýkoľvek iný):

Napíšte prosím príklad pripojenia k MySQL a konverzie údajov do ↴

💡 Odpoveď LLM :

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7     user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Run in IDE

Obrázok 7.2-6 Príklad pripojenia cez Python k databáze MySQL a importovanie údajov z databázy MySQL do DataFrame .

Výsledný kód (obr. 7.2-5, obr. 7.2-6) je možné spustiť v jednom z populárnych IDE (integrovaných vývojových prostredí), ktoré sme spomenuli vyššie, v offline režime: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo populárne online nástroje: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Načítaním multiformátových údajov do premennej "df" (obrázok 7.2-5 - riadok 25; obrázok 7.2-6 - riadok 8) sme údaje previedli do formátu Pandas DataFrame, jednej z najpopulárnejších štruktúr na spracovanie údajov, čo je dvojrozmerná tabuľka s riadkami a stĺpcami. O ďalších formátoch ukladania údajov používaných v ETL -Pipeline, ako sú Parquet, Apache ORC, JSON, Feather, HDF5, ako aj moderné dátové sklady, budeme viac hovoriť v kapitole "Ukladanie a správa údajov v stavebníctve" (obrázok 8.1-2).

Po fáze extrakcie a štruktúrovania údajov (Extract) sa vytvorí jedno pole informácií (obr. 7.2-5, obr. 7.2-6), ktoré je pripravené na ďalšie spracovanie. Pred vložením týchto údajov do cieľových systémov alebo ich použitím na analýzu je však potrebné zabezpečiť ich kvalitu, integritu a súlad so stanovenými požiadavkami. V tejto fáze dochádza k transformácii údajov (Transformácia) - kľúčový krok na zabezpečenie spoľahlivosti následných záverov a rozhodnutí.

## ETL Transformácia: aplikácia pravidiel validácie a transformácie

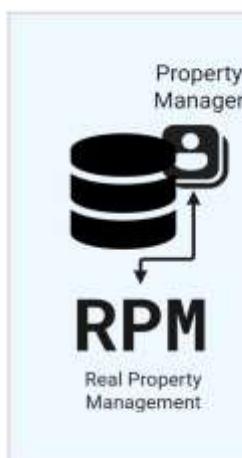
V kroku Transformácia sa údaje spracujú a transformujú. Tento proces môže zahŕňať kontrolu správnosti, normalizáciu, doplnenie chýbajúcich hodnôt a validáciu pomocou automatizovaných nástrojov

Podľa štúdie spoločnosti PwC "Data-Driven. What Students Need to Succeed in a Rapidly Changing Business World" (2015) [9] sa moderné audítorské spoločnosti odkláňajú od náhodného overovania údajov a prechádzajú na analýzu obrovského množstva informácií pomocou automatizovaných nástrojov. Tento prístup umožňuje nielen identifikovať nezrovnalosti vo výkazníctve, ale aj ponúkať odporúčania na optimalizáciu podnikových procesov.

V stavebníctve sa podobné metódy môžu používať napríklad na automatickú validáciu údajov o návrhu, kontrole kvality stavby a hodnotenie výkonnosti dodávateľa. Jedným z nástrojov na automatizáciu a urýchlenie spracovania údajov je používanie regulárnych výrazov (RegEx) vo fáze transformácie údajov (Transform) procesu ETL. RegEx umožňuje efektívne overovať reťazce údajov, zisťovať nekonzistentnosti a zabezpečovať integritu informácií s minimálnou spotrebou zdrojov. Podrobnejšie informácie o RegEx (obr. 4.4-7) boli opísané v kapitole "Preklad požiadaviek do štruktúrovanej podoby".

Uvedme si praktický príklad: v systéme správy objektov majetku (RPM) manažér nastavuje požiadavky na kľúčové atribúty objektov (obr. 7.2-7). Vo fáze transformácie je potrebné overiť nasledujúce parametre:

- overenie formátov identifikátorov objektov (atribút "ID")
- kontrola hodnôt náhradnej záručnej doby (atribút "Záručná doba")
- overenie cyklu výmeny prvkov (atribút "Požiadavky na údržbu")



**Property Manager:  
Long-term Management**

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-OLD1	Window	8 years	15 years	Biannual Inspection
W-OLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Maintenance

Obrázok 7.2-7 Overovanie kvality začína stanovením požiadaviek na atribúty a ich hraničných hodnôt.

Na nastavenie hraničných hodnôt pre overovanie parametrov napríklad predpokladajme, že z našich skúseností vieme, že platné hodnoty pre atribút "ID" môžu obsahovať len reťazcové hodnoty "W-NEW",

"W-OLD1" alebo "D-122" alebo podobné hodnoty, kde prvým znakom je písmeno, za ktorým nasleduje pomlčka a potom tri abecedné znaky "NEW", "OLD" alebo akékoľvek trojmiestne číslo (obrázok 7.2-7). Na overenie týchto identifikátorov možno použiť nasledujúci regulárny výraz (RegEx):

```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Táto šablóna zabezpečuje, aby všetky identifikátory v údajoch zodpovedali zadaným kritériám. Ak niektorá hodnota nevyhovie kontrole, systém zaznamená chybu. Ak chcete vytvoriť Python -kód na transformáciu údajov a použiť výsledné údaje na vytvorenie tabuľky výsledkov, stačí sformulovať dotaz v rozhovore LLM.

💡 Textová požiadavka na LLM:

Napište kód na overenie stĺpcov DataFrame pomocou regulárnych výrazov, ktorý kontroluje identifikátory vo formáte "W-NEW" alebo "W-OLD" pomocou RegEx, energetickú účinnosť s písmenami "A" až "G", záručnú dobu a cyklus výmeny s číselnými hodnotami v rokoch ↴

💡 Odpoveď LLM :

```
1 # RegEx
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r"^(W-NEW|W-OLD[0-9]+|D-1[0-9]{2})$",
7     'Energy_Efficiency': r"^[A-G]\w*$",
8     'Warranty_Period': r"\d+\w+years$",
9     'Replacement_Cycle': r"\d+\w+years$"
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20
```

Obrázok 7.2-8 Kód automatizuje proces validácie použitím šablón RegEx na stĺpce parametrov dátového rámca .

Vyššie uvedený, automaticky generovaný kód jazyka Python (obrázok 7.2-8), používa knižnicu "re" (regulárne výrazy RegEx) na definovanie funkcie, ktorá kontroluje každý atribút dátovej položky v DataFrame. Pre každý zadaný stĺpec (atribút) funkcia použije vzor RegEx na kontrolu, či každý záznam zodpovedá očakávanému formátu, a výsledky pridá ako nové hodnoty (False/True) do nového atribútu-stĺpca DataFrame.

Takáto automatizovaná validácia zabezpečuje formálnu zhodu údajov a môže sa použiť ako súčasť systému kontroly kvality vo fáze transformácie.

Po úspešnom dokončení kroku transformácie a kontroly kvality sú údaje pripravené na odoslanie do cieľových systémov. Transformované a overené údaje možno nahrať do CSV, JSON, Excel, databáz a iných formátov na ďalšie použitie. V závislosti od úlohy možno výsledky prezentovať aj v správach, grafoch alebo analytických paneloch.

## ETL Load: Vizualizujte výsledky v tabuľkách a grafoch

Po ukončení etapy Transform, keď sú údaje štruktúrované a overené, nasleduje posledná etapa Load, v ktorej sa údaje môžu načítať do cieľového systému a zároveň sa môžu vizualizovať na účely analýzy. Vizuálna prezentácia údajov umožňuje rýchlo identifikovať odchýlky, analyzovať rozdelenia a označiť klúčové závery všetkým účastníkom projektu vrátane tých, ktorí nemajú technické vzdelenie.

Namiesto prezentácie informácií vo forme tabuľiek a čísel môžeme použiť infografiky, grafy a informačné panely (dashboards). Jedným z najbežnejších a najflexibilnejších nástrojov na vizualizáciu štruktúrovaných údajov v jazyku Python je knižnica Matplotlib (obrázok 7.2-9, obrázok 7.2-10). Umožňuje vytvárať statické, animované a interaktívne grafy a podporuje širokú škálu typov grafov.

- Na vizualizáciu výsledkov kontroly atribútov zo systému RPM (obrázok 7.2-7), môžete použiť nasledujúci dotaz na jazykový model:

Napíšte kód na vizualizáciu údajov DataFrame, vyššie (obrázok 7.2-7), s histogramom pre výsledky, ktorý zobrazí frekvenciu chýb v atribúte ↵

- ❸ Odpoveď LLM vo forme kódu a pripravená vizualizácia výsledkov vykonávania kódu priamo v chate LLM :

```

Create in LLM Chat
Pipeline.py
Run in IDE

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

```

**Data Verification Summary - Bar Chart**

Data Category	Passed (Green)	Failed (Red)	Total Count
Window_ID	4.0	1.0	5.0
Energy_Efficiency	3.0	2.0	5.0
Warranty_Period	4.0	1.0	5.0
Replacement_Cycle	4.0	1.0	5.0

Obr. 7.2-9 Vizualizácia výsledkov kroku Transformácia kontroly hodnôt atribútov zo systému RPM (obr. 7.2-7) ako histogram v kroku Načítanie .

- 2 Existuje mnoho bezplatných knižníc s otvoreným zdrojovým kódom, ktoré umožňujú reprezentovať štruktúrované údaje v rôznych formátoch. Pokračujme vo vizualizácii výsledkov pomocou iného typu grafu s nasledujúcou výzvou v chate:

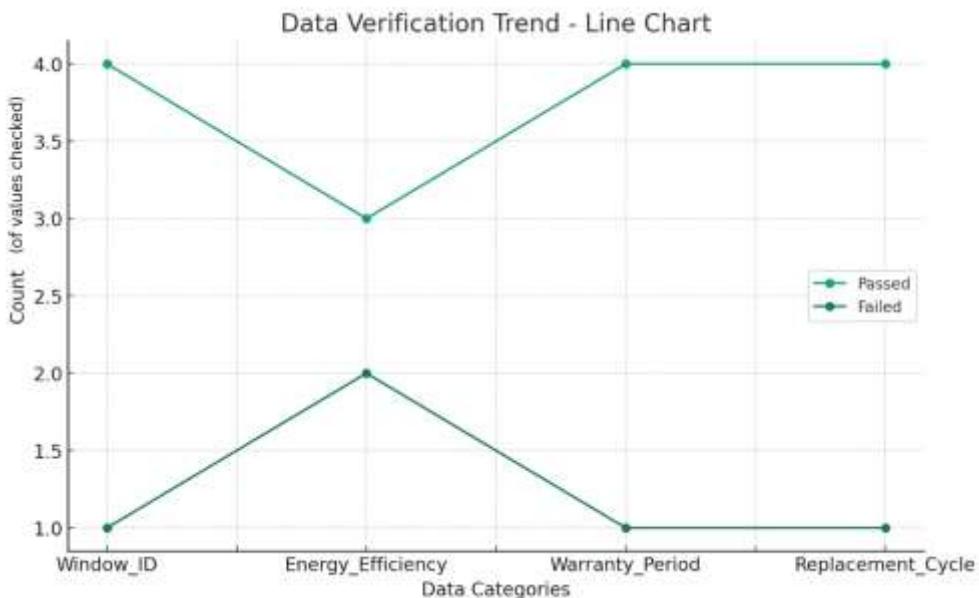
Zakreslite tie isté údaje ako graf čiar ↵

💡 Odpoveď LLM :

```

1 # Visualization 3: Line Chart
2 df_visual.plot(kind='line', marker='o', figsize=(10, 6))
3 plt.title('Data Verification Trend - Line Chart')
4 plt.xlabel('Data Categories')
5 plt.ylabel('Count')
6 plt.grid(True)
7 plt.show()

```



Obrázok 7.2-10 Vizualizácia validačných údajov (obrázok 7.2-8) ako čiarový diagram získaný pomocou knižnice Matplotlib .

Existuje mnoho bezplatných knižníc s otvoreným zdrojovým kódom pre vizualizáciu, ako napr.:

- Seaborn - pre štatistické grafy (obr. 7.2-11)
- Plotly - pre interaktívne webové vizualizácie (obr. 7.2-12, obr. 7.1-6)
- Altair - na deklaratívnu vizualizáciu
- Dash alebo Streamlit - na vytvorenie úplných informačných panelov

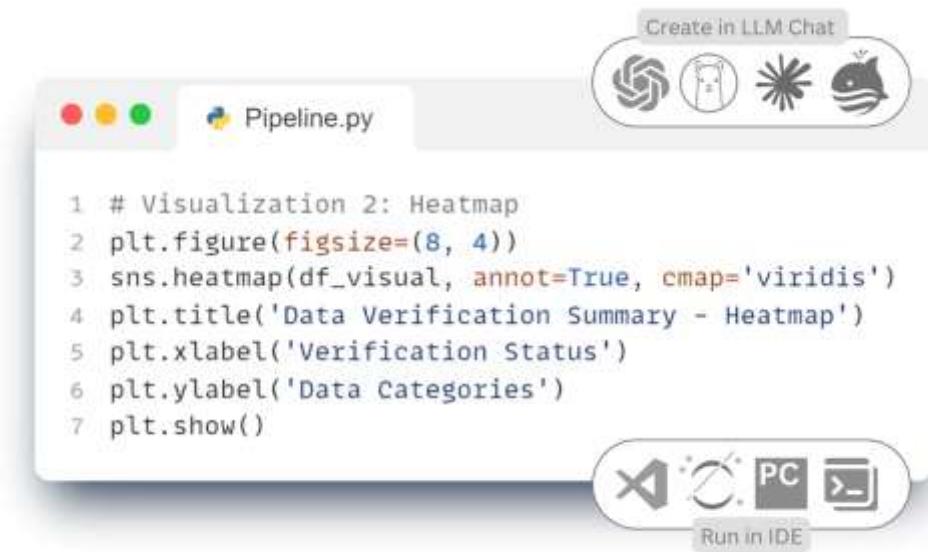
Znalosť konkrétnych knižníc na vizualizáciu nie je potrebná - moderné nástroje vrátane LLM, umožňujú automaticky generovať kód grafov a celých aplikácií na základe opisu úlohy.

Výber nástroja závisí od cieľov projektu: či ide o správu, prezentáciu alebo online informačný panel. Napríklad open source knižnica Seaborn je obzvlášť vhodná na prácu s kategorickými údajmi, pomáha identifikovať vzory a trendy.

- Ak chcete vidieť fungujúcu knižnicu Seaborn, môžete požiadať o používanie požadovanej knižnice priamo LLM alebo poslať podobnú textovú žiadosť LLM na adresu:

Zobrazenie tepelnej mapy ↗

- Odpoveď LLM vo forme kódu a hotového grafu, ktorého kód na vykreslovanie možno teraz skopírovať do IDE a samotný graf možno skopírovať alebo uložiť na vloženie do dokumentu:



The screenshot shows a code editor window with the following Python code:

```

1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

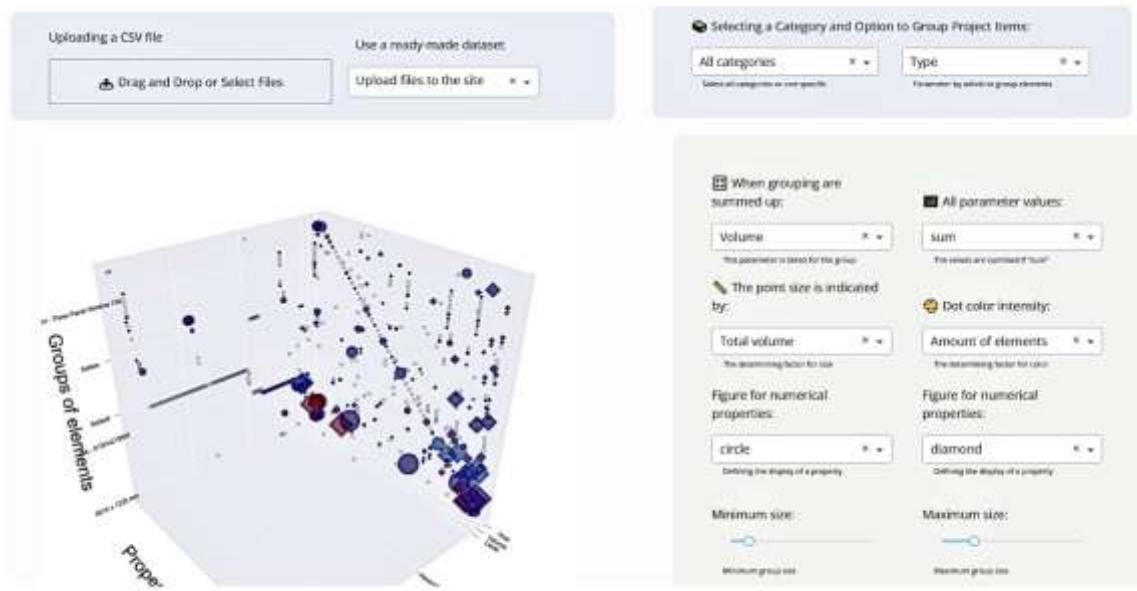
```

The interface includes a toolbar at the top with icons for file operations and a "Create in LLM Chat" button. Below the code editor, there are icons for running the code in various environments: VS Code, Jupyter Notebook, PC, and a terminal. A "Run in IDE" button is also present.



Obrázok 7.2-11 Vizualizácia výsledkov validácie (obrázok 7.2-8) údajov pomocou knižnice Seaborn.

Pre tých, ktorí uprednostňujú interaktívny prístup, existujú nástroje, ktoré umožňujú vytvárať dynamické grafy a panely s možnosťou interakcie. Knižnica Plotly (obrázok 7.1-6, obrázok 7.2-12) ponúka možnosť vytvárať vysoko interaktívne grafy a panely, ktoré možno vložiť do webových stránok a umožniť používateľovi interakciu s údajmi v reálnom čase.



Obr. 7.2-12 Interaktívna 3D -vizualizácia atribútov prvkov z projektu CAD- (BIM-) pomocou knižnice Plotly.

Špecializované knižnice s otvoreným zdrojovým kódom Bokeh, Dash a Streamlit poskytujú pohodlný spôsob prezentácie údajov bez potreby hlbokých znalostí vývoja webových stránok. Bokeh je vhodný na tvorbu komplexných interaktívnych grafov, Dash sa používa na vytváranie plnohodnotných analytických panelov a Streamlit umožňuje rýchlo vytvárať webové aplikácie na analýzu údajov.

Pomocou takýchto vizualizačných nástrojov môžu vývojári a analytici efektívne šíriť výsledky medzi kolegami a zainteresovanými stranami, čo umožňuje intuitívnu interakciu s údajmi a zjednodušuje rozhodovanie.

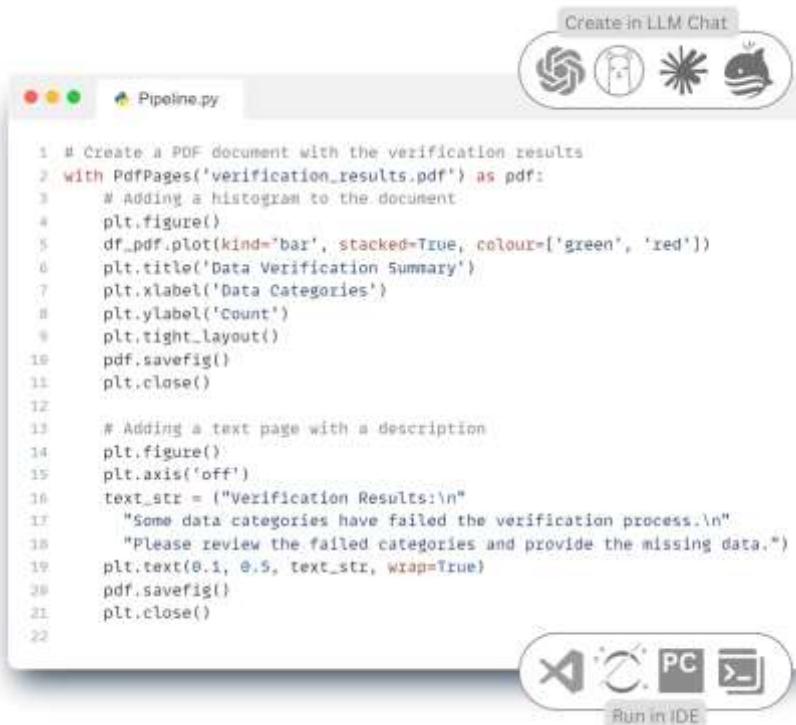
## ETL Load: Automatické vytváranie dokumentov PDF

Vo fáze načítania údajov môžete nielen vizualizovať údaje, nahrávať ich do tabuľiek alebo databáz, ale aj automaticky generovať správy vrátane potrebných grafov, diagramov a kľúčových analytických ukazovateľov, ktoré dostane manažér alebo odborník čakajúci na výsledky auditu. Automatizované správy môžu obsahovať komentáre aj textovú interpretáciu údajov a vizualizačné prvky - tabuľky, grafy.

- Na vytvorenie PDF -správy s histogramom (obr. 7.2-9) a opisom analýzy na základe validácie, ktorú sme vykonali v predchádzajúcich kapitolách, stačí sformulovať požiadavku v pokračovaní dialógu s LLM, napr.:

Napište kód na vytvorenie súboru PDF -file s histogramom a popisom výsledkov validácie vyššie uvedených údajov (v chate) a napište textové upozornenie, že niektoré kategórie neboli validované a že je potrebné doplniť chýbajúce ↵

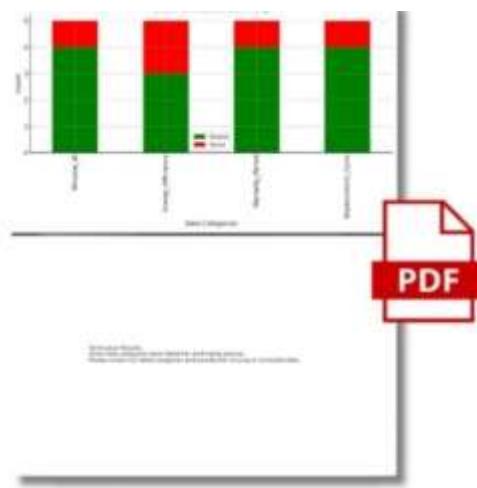
- Odpoveď LLM ako kód a hotové PDF s výsledkami:



```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # Adding a histogram to the document
4     plt.figure()
5     df.pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "Please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



**Obrázok 7.2-13 Automatický kód vytvorí PDF -dokument obsahujúci histogram s testovacími údajmi a text s výsledkami testu.**

Automaticky napísané riešenie pozostávajúce len z 20 riadkov kódu pomocou LLM okamžite vytvorí požadovaný dokument PDF (alebo DOC) s vizualizáciou vo forme histogramu atribútov (obr. 7.2-13) zobrazujúceho počet údajov, ktoré prešli a neprešli validáciou, a s pridaním textového bloku so zhrnutím výsledkov a odporúčaniami ďalší postup.

Automatizované generovanie dokumentov je kľúčovým prvkom fázy Load, najmä v projektovom prostredí, kde je rozhodujúca rýchlosť vykazovania a presnosť.

## ETL Load: automatické generovanie dokumentov z FPDF

Automatizácia reportovania vo fáze ETL Load je dôležitým krokom pri spracovaní údajov, najmä ak je potrebné prezentovať výsledky analýz vo formáte, ktorý je ľahko komunikovateľný a zrozumiteľný. V stavebníctve je to často relevantné pre správy o priebehu prác, štatistiky údajov o projekte, správy o zabezpečení kvality alebo finančnú dokumentáciu.

Jedným z najvhodnejších nástrojov na takéto úlohy je open source knižnica, FPDF, dostupná pre Python aj PHP.

Knižnica **FPDF** s otvoreným zdrojovým kódom poskytuje flexibilný spôsob generovania dokumentov prostredníctvom kódu, ktorý umožňuje pridať hlavičky, text, tabuľky a obrázky. Používanie kódu namiesto ručných úprav znižuje počet chýb a urýchľuje proces prípravy správ vo formáte PDF.

Jedným z kľúčových krokov pri vytváraní dokumentu PDF je pridanie nadpisov a hlavného textu vo forme komentárov alebo popisov. Pri vytváraní správy je však dôležité nielen pridať text, ale aj správne ho štruktúrovať. Nadpisy, odrážky, riadkovanie - to všetko ovplyvňuje čitateľnosť dokumentu. Pomocou FPDF, môžete nastaviť parametre formátovania, kontrolovať usporiadanie prvkov a prispôsobiť štýl dokumentu.

FPDF je v princípe veľmi podobný HTML. Tí, ktorí už poznajú HTML, môžu pomocou FPDF ľahko vytvárať dokumenty PDF akejkoľvek zložitosti, pretože štruktúra kódu je veľmi podobná značkovaniu HTML: hlavičky, text, obrázky a tabuľky sa pridávajú podobným spôsobom. Tí, ktorí nepoznajú jazyk HTML, sa nemusia obávať - môžete použiť LLM, ktorý vám okamžite pomôže zostaviť kód na vygenerovanie požadovaného rozloženia dokumentu.

- Nasledujúci príklad ukazuje, ako vygenerovať správu so záhlavím a textom tela. Spustením tohto kódu v ľubovoľnom IDE s podporou jazyka Python sa vytvorí súbor PDF -file obsahujúci požadované záhlavie a text:

```
from fpdf import FPDF      # Import knižnice FPDF
pdf = FPDF()    # Vytvorte PDF -dokument
pdf.add_page()  # Pridať stránku

pdf.set_font("Arial", style='B', size=16) # Nastavenie písma: Arial, tučné, veľkosť 16
pdf.cell(200, 10, "Project Report", ln=True, align='C') # Vytvorte nadpis a zarovnajte ho na stred
pdf.set_font("Arial", size=12) # Zmeňte písmo na obyčajné Arial, veľkosť 12
pdf.multi_cell(0, 10, "Tento dokument obsahuje údaje o výsledkoch overovania projektových súborov...") # Pridajte viacriadkový text
pdf.output(r "C:\reports\report.pdf") # Uložiť PDF -file
```



**Obrázok 7.2-14** Pomocou niekoľkých riadkov kódu jazyka Python môžeme automaticky vygenerovať textový dokument PDF, ktorý potrebujeme.

Pri príprave správ je dôležité vziať do úvahy, že údaje, z ktorých sa dokument tvorí, sú len zriedka statické. Hlavičky, textové bloky (obr. 7.2-14) sú často tvorené dynamicky, dostávajú hodnoty vo fáze Transformácia v procese ETL.

Použitie kódu umožňuje vytvárať dokumenty, ktoré obsahujú aktuálne informácie: názov projektu, dátum vytvorenia správy, ako aj informácie o účastníkoch alebo aktuálnom stave. Použitie premenných v kóde umožňuje automaticky vkladať tieto údaje na požadované miesta v správe, čím sa úplne eliminuje potreba ručnej úpravy pred odoslaním.

Okrem jednoduchého textu a nadpisov majú v projektovej dokumentácii osobitné miesto tabuľky. Takmer každý dokument obsahuje štruktúrované údaje: od popisov objektov až po výsledky kontrol.

Automatické generovanie tabuľiek na základe údajov z etapy transformácie umožňuje nielen urýchliť proces prípravy dokumentu, ale aj minimalizovať chyby pri prenose informácií. FPDF umožňuje vkladať tabuľky do PDF - súborov (ako text alebo obrázky), nastavovať okraje buniek, veľkosť stĺpcov a písma (obr. 7.2-15). Je to výhodné najmä pri práci s dynamickými údajmi, keď sa počet riadkov a stĺpcov môže meniť v závislosti od úloh dokumentu.

- Nasledujúci príklad ukazuje, ako automatizovať vytváranie tabuľiek, napr. so zoznamami materiálov, odhadmi alebo výsledkami testov parametrov:

```
údaje = [
    [ "Položka", "Množstvo", "Cena"],   # Nadpisy stĺpcov
    [ "Betón", "10 m3", "$ 500."],   # Údaje prvého riadku
    [ "Rebar", "2 tony", "$ 600"],      # Údaje druhého riadku .
    [ "Tehla", "5 000 kusov", "$ 750."], # Údaje tretieho riadku .
]

pdf = FPDF () # Vytvorenie PDF -dokument
pdf.add_page() # Pridanie stránky
pdf.set_font("Arial", size=12) # Nastavenie písma

for row in data: # Vyhľadávanie riadkov tabuľky
    for item in row: # Prejdite bunky v riadku
        pdf.cell(60, 10, item, border=1) # Vytvorte bunku s ohriadením, šírkou 60 a výškou 10
    pdf.ln() # Presun na ďalší riadok
pdf.output(r "C:\reports\table.pdf") # Uložiť PDF -file
```

Item	Quantity	Price
Concrete	10 m <sup>3</sup>	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

Obrázok 7.2-15 Z kroku Transformácia môžete automaticky generovať nielen text, ale aj akékoľvek informácie o tabuľke v PDF.

V reálnych scenároch vykazovania sú tabuľky zvyčajne dynamicky generované informácie získané vo fáze transformácie údajov. V uvedenom príklade (obr. 7.2-15) je tabuľka vložená do PDF -dokumentu v statickej podobe: údaje pre príklad boli umiestnené v dátovom slovníku (prvý riadok kódu), v reálnych podmienkach sa takáto premenná údajov doplní automaticky napr. po zoskupení dátového rámca.

V praxi sa takéto tabuľky často vytvárajú na základe štruktúrovaných údajov pochádzajúcich z rôznych dynamických zdrojov: databáz, súborov Excel, rozhraní API alebo výsledkov analytických výpočtov. Najčastejšie sa vo fáze transformácie (ETL) údaje agregujú, zoskupujú alebo filtrovú - a až potom sa transformujú do súhrnných údajov vo forme grafov alebo dvojrozmerných tabuľiek zobrazovaných v prehľadoch. To znamená, že obsah tabuľky sa môže meniť v závislosti od zvolených parametrov,

obdobia analýzy, filtrov projektu alebo nastavení používateľa.

Vďaka použitiu dynamických dátových rámcov a súborov údajov vo fáze Transform je proces vykazovania vo fáze Load čo najflexibilnejší, škálovateľný a ľahko opakovateľný bez potreby manuálneho zásahu.

Okrem tabuľiek a textu FPDF podporuje aj pridávanie grafov tabuľkových údajov, čo umožňuje vložiť do správy obrázky vytvorené pomocou Matplotlib alebo iných vizualizačných knižníc, o ktorých sme hovorili vyššie. Akékolvek grafy, diagramy a schémy môžete do dokumentu pridať pomocou kódu.

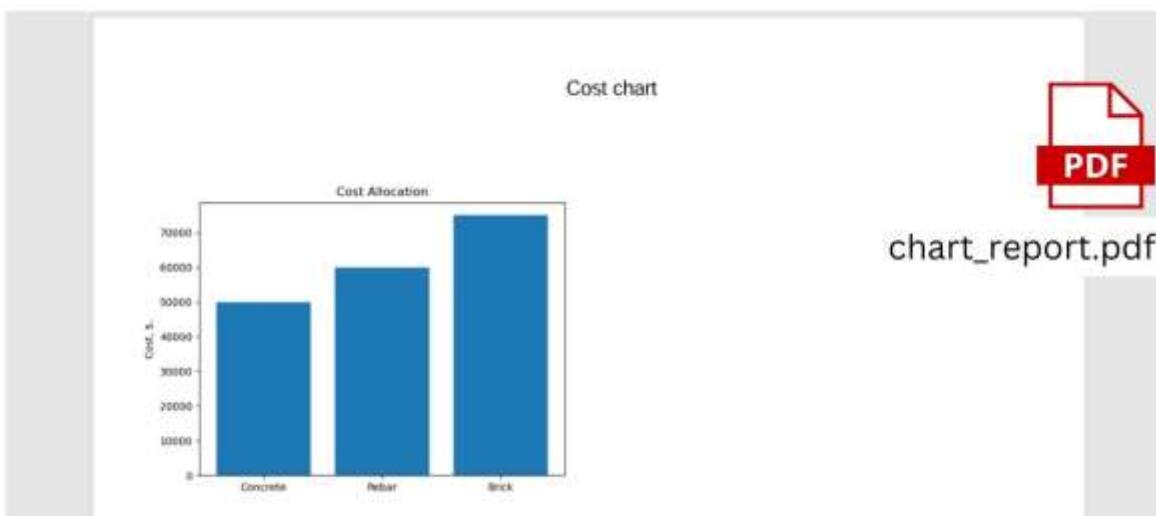
- Pomocou knižnice Python FPDF, pridáme do dokumentu PDF graf vopred vygenerovaný pomocou Matplotlib:

```
import matplotlib.pyplot as plt # Import Matplotlib na vytvorenie grafov

fig, ax = plt.subplots() # Vytvorte obrázok a osi grafu
categories = ["Concrete", "Rebar", "Brick"] # Názvy kategórií
values = [50000, 60000, 75000] # Hodnoty kategórií
ax.bar(categories, values) # Vytvorte stípcový graf
plt.ylabel("Hodnota,$.") # Označte os Y
plt.title("Rozdelenie nákladov") # Pridajte nadpis
plt.savefig(r"C:\reports\chart\chart\chart.png") # Uložte graf ako obrázok

pdf = FPDF () # Vytvorte PDF -dokument
pdf.add_page() # Pridať stránku
pdf.set_font("Arial", size=12) # Nastavte písmo
pdf.cell(200, 10, "Cost Chart", ln=True, align="C") # Pridať záhlavie

pdf.image(r"C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Vložte obrázok do
# PDF (x, y - súradnice, w - šírka)
pdf.output(r"C:\reports\chart_report.pdf") # Uložiť súbor PDF
```



Obr. 7.2-16 Pomocou desiatok riadkov kódu môžete vygenerovať graf, uložiť ho a potom vložiť do dokumentu PDF.

S FPDF sa proces prípravy dokumentov a logiky stáva transparentným, rýchlym a pohodlným. Šablóny zabudované v kóde umožňujú generovať dokumenty s aktuálnymi údajmi, čím sa eliminuje potreba ručného vypĺňania.

Používanie automatizácie ETL - namiesto časovo náročného manuálneho reportovania sa odborníci môžu sústrediť na analýzu údajov a prijímanie rozhodnutí, a nie na výber správneho nástroja na prácu s konkrétnym dátovým silom s prehľadným používateľským rozhraním.

Knižnica FPDF tak poskytuje flexibilný nástroj na automatizovanú tvorbu dokumentov akejkoľvek zložitosti - od krátkych technických správ až po komplexné analytické súhrny s tabuľkami a grafmi, čo umožňuje nielen urýchliť tok dokumentov, ale aj výrazne znížiť pravdepodobnosť chýb spojených s ručným zadávaním a formátovaním údajov.

## ETL Načítanie: Hlásenie a načítanie do iných systémov

Vo fáze Load sa výsledky generovali vo forme tabuľiek, grafov a záverečných správ vo formáte PDF pripravených v súlade so stanovenými požiadavkami. Ďalej je možné tieto údaje exportovať do strojovo čitateľných formátov (napr. CSV), čo je potrebné na integráciu s externými systémami, ako sú ERP, CAFM, CPM, platformy BI a iné podnikové alebo odvetvovové riešenia. Okrem formátu CSV je možné odosielať údaje do formátov XLSX, JSON, XML alebo priamo do databáz, ktoré podporujú automatickú výmenu informácií.

- 💡 Ak chcete vygenerovať príslušný kód na automatizáciu kroku Load, jednoducho sa opýtajte napríklad na rozhranie LLM -interface: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude alebo QWEN:

Napište kód na generovanie správy o výsledkoch validácie údajov v DataFrame, kde sa spočítajú stĺpce s prefixom "verified\_ ", premenujú sa na "Passed" a "Failed", chýbajúce hodnoty sa nahradia hodnotou 0 a potom sa do súboru CSV vyexportujú len tie riadky, ktoré prešli všetkými validáciami.

#### Odpoveď LLM:



```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts());T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

Obrázok 7.2-17 Overené údaje získané v kroku Transformácia z konečného rámca údajov sa exportujú do súboru CSV - na integráciu s inými systémami.

Vyššie uvedený kód (obr. 7.2-17) implementuje záverečnú fázu procesu ETL -process - Load, počas ktorej sa overené údaje uložia vo formáte CSV, kompatibilnom s väčšinou externých systémov a databáz. Tým sme ukončili celý cyklus procesu ETL -procesu, vrátane extrakcie, transformácie, vizualizácie, dokumentácie a exportu údajov do potrebných systémov a formátov, čo zabezpečuje reproducibilnosť, transparentnosť a automatizáciu práce s informáciami.

Potrubie ETL možno použiť na spracovanie jednotlivých projektov aj na rozsiahle aplikácie - pri analýze stoviek alebo tisícov prichádzajúcich údajov vo forme dokumentov, obrázkov, skenov, projektov CAD, mrakov bodov, súborov PDF alebo iných zdrojov pochádzajúcich z distribuovaných systémov. Schopnosť plne automatizovať proces robí z ETL nielen technický nástroj na spracovanie, ale aj základ digitálnej informačnej infraštruktúry v stavebníctve.

## ETL s LLM: Vizualizácia údajov z PDF -dokumentov

Je čas prejsť na vytvorenie plnohodnotného procesu ETL, ktorý pokrýva všetky kľúčové fázy spracovania údajov v jednom scenári - extrakciu, transformáciu a načítanie. Vytvorme automatizovaný ETL-Pipeline, ktorý vám umožní spracovať dokumenty PDF bez manuálnej práce - extrahovať údaje z dokumentov, vizualizovať, analyzovať a prenášať do iných systémov.

Proces ETL v našom príklade bude opísaný prostredníctvom výziev, ktoré budú musieť vysvetliť

jazykovému modelu (LLM) všetky procesy ETL s opisom konečného výsledku, ktorý sa má získať. V tomto prípade je úlohou nájsť všetky súbory PDF v zadanom priečinku a jeho podpriečinkoch, extrahovať z nich relevantné informácie - napr. názvy materiálov, množstvá a náklady - a prezentovať výsledok ako štruktúrovanú tabuľku (DataFrame) na ďalšiu analýzu

- ➊ Prvý textový dotaz v LLM na automatické extrahovanie údajov z viacerých dokumentov PDF a vytvorenie dátového rámca údajov pre krok Extract:

Napište kód na extrahovanie informácií o materiáloch zo súborov PDF v danom priečinku a jeho podpriečinkoch. Údaje v PDF obsahujú názov materiálu, množstvo a cenu. Výsledok by mal byť uložený v DataFrame ↴

- ➋ Odpoveď LLM :

```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26
27    # Convert list of dictionaries to pandas DataFrame
28    return pd.DataFrame(data)
29
30 # Specify the base folder path containing PDF files and subfolders with PDFs
31 df = extract_data('C:\\')

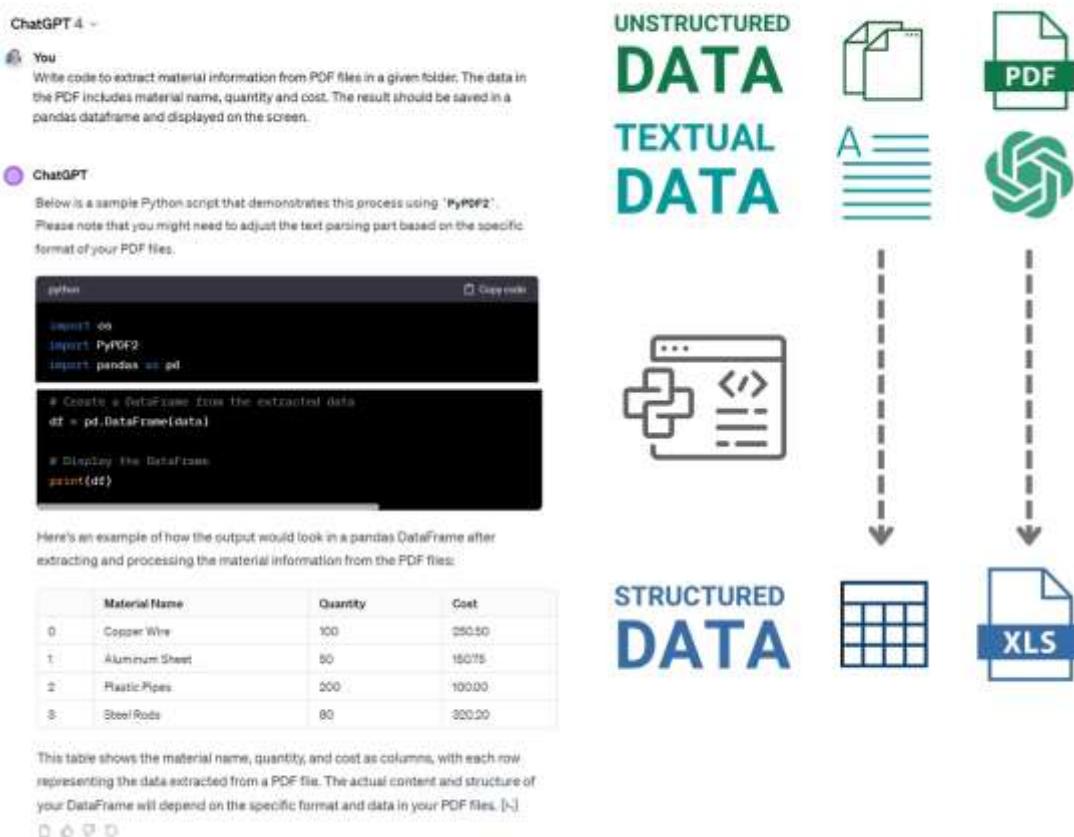
```

Obrázok 7.2-18 LLM vytvorí kód jazyka Python -kód na extrakciu údajov zo súborov PDF -v konkrétnom priečinku a všetkých jeho podpriečinkoch.

Odpoveď LLM (obrázok 7.2-18) je hotový skript v jazyku Python, ktorý automaticky prechádza všetky

priečinky, otvára nájdené súbory PDF, extrahuje z nich textové informácie a konvertuje ich do tabuľky. Výsledný chatový kód možno spustiť v niektorom z populárnych IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo v populárnych online nástrojoch Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Alternatívne, aby sme zjednodušili proces zberu, namiesto kopírovania kódu z LLM a používania kódu v IDE môžeme tiež nahrať tucet PDF - súborov priamo do chatu LLM (obrázok 7.2-19) a získať tabuľku ako výstup bez toho, aby sme museli vidieť kód alebo ho spustiť. Výsledkom vykonania tohto kódu bude tabuľka s atribútmi, ktoré sme vybrali.

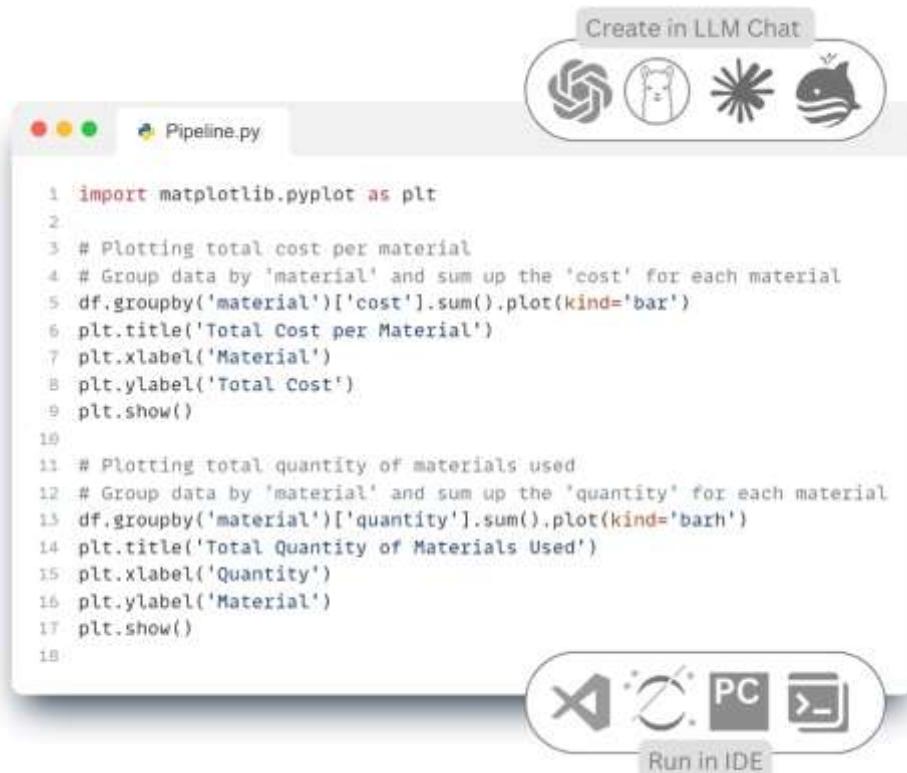


Obrázok 7.2-19 Výsledok vykonania kódu v LLM, ktorý extrahuje údaje zo súborov PDF - v štruktúrovanom zobrazení dátového rámca s vybranými atribútmi.

V ďalšom kroku požiadame o jazykový model na získaných údajoch - napríklad na porovnanie nákladov a objemu spotreby materiálu a vytvorenie niekoľkých príkladov vizualizácií, ktoré budú slúžiť ako základ pre ďalšiu analýzu.

- ❷ Požiadajte v pokračujúcim rozhovore s LLM o vykreslenie niektorých grafov z tabuľiek, ktoré boli vytvorené v kroku Transformácia (obrázok 7.2-18):

Vizualizujte celkové náklady a množstvo každého materiálu z DataFrame (obrázok 7.2-18)



The screenshot shows a chat interface with a message box containing Python code. The code uses the matplotlib library to plot total cost and quantity for five materials: Wood, Steel, Concrete, Glass, and Plastic. The first plot is a bar chart titled 'Total Cost per Material' with 'Material' on the x-axis and 'Total Cost' on the y-axis. The second plot is a bar chart titled 'Total Quantity of Materials Used' with 'Material' on the x-axis and 'Quantity' on the y-axis. A pie chart titled 'Cost Distribution by Material' shows the percentage distribution of costs among Steel, Wood, Glass, and Plastic.

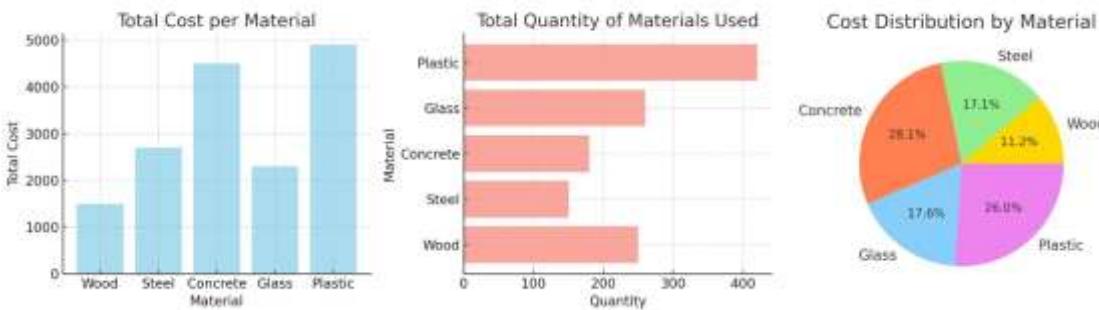
```

1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

Obrázok 7.2-20 Odozva LLM -modelu ako kód jazyka Python na vizualizáciu údajov z dátového rámca pomocou knižnice matplotlib.

LLM automaticky generuje a vykonáva Python -kód (obr. 7.2-20) pomocou knižnice matplotlib. Po spustení tohto kódu získame grafy nákladov a spotreby materiálu v stavebných projektoch priamo v chate (obr. 7.2-21), čo výrazne zjednodušuje analytickú prácu.

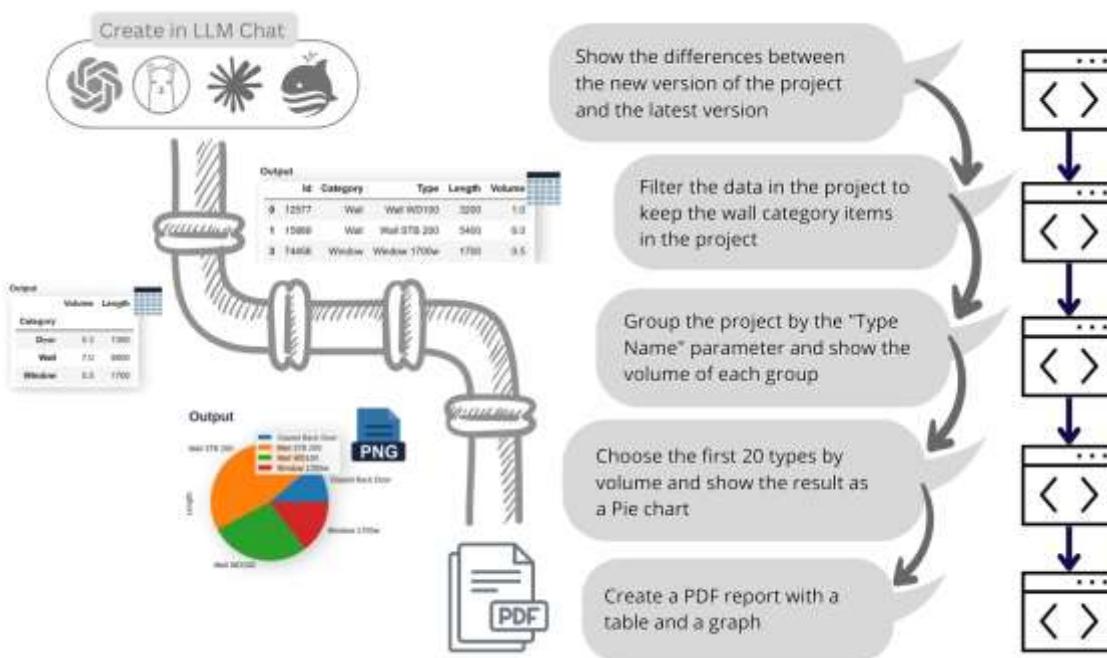


Obrázok 7.2-21 Vizualizácia odozvy LLM vo forme grafov na základe údajov zhromaždených v DataFrame.

Podpora pri tvorbe nápadov na písanie kódu ETL, analýza a vykonávanie kódu a vizualizácia výsledkov

je k dispozícii prostredníctvom jednoduchých textových dotazov v LLM, bez potreby učiť sa základy programovania. Vznik nástrojov umelej inteligencie, ako je LLM, definitívne mení prístup k programovaniu a automatizácii spracovania údajov (obrázok 7.2-22).

Podľa správy spoločnosti PwC "Aká je skutočná hodnota umelej inteligencie pre váš podnik a ako ju môžete využiť?" (2017) [139] budú hlavnými hnacími silami hospodárskeho rastu automatizácia procesov a zvyšovanie produktivity. A očakáva sa, že zlepšenie produktivity bude v rokoch 2017 až 2030 predstavovať viac ako 55% všetkého rastu HDP poháňaného umelou inteligenciou."



Obrázok 7.2-22 AI LLM pomáha vytvárať návrh kódu, ktorý sa použije v budúcich projektoch bez potreby LLM.

Pomocou nástrojov, ako sú ChatGPT, LlaMa, Mistral, Claude, DeepSeek, QWEN, Grok, ako aj otvorených údajov a softvéru s otvoreným zdrojovým kódom, môžeme automatizovať procesy, ktoré sa predtým vykonávali len pomocou špecializovaných, nákladných a ťažko udržiavateľných modulárnych proprietárnych systémov.

V kontexte stavebníctva to znamená, že spoločnosti, ktoré ako prvé zavedú automatizované procesy Pipeline -data, získajú významné výhody, od vyšej efektívnosti riadenia projektov cez zníženie finančných strát až po odstránenie roztrieštených aplikácií a izolovaných dátových skladov.

Popísaná logika vykonávania obchodných úloh v procese ETL je dôležitou súčasťou automatizácie procesov analýzy a spracovania údajov, ktorá je špecifickou variáciou širšieho pojmu - potrubia (Pipelines).

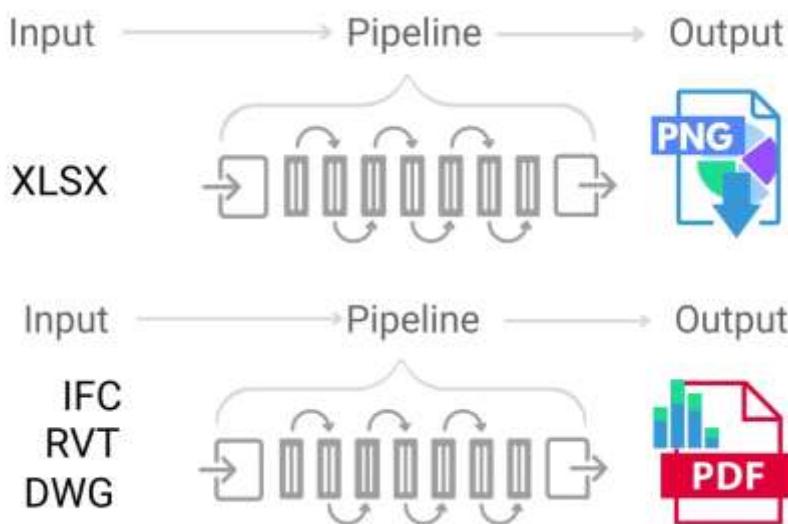
## KAPITOLA 7.3.

### AUTOMATICKÝ ETL DOPRAVNÍK (POTRUBIE)

#### Potrubie: Automatický dopravník ETL údaje

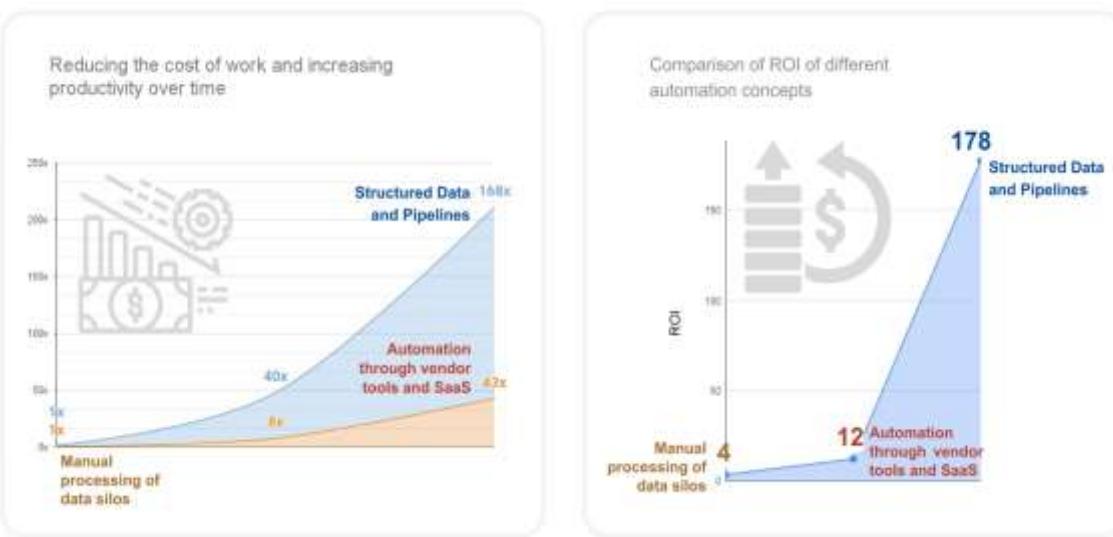
Proces ETL sa tradične používa na spracovanie údajov v analytických systémoch, pričom zahŕňa štruktúrované aj neštruktúrované zdroje. V dnešnom digitálnom prostredí sa však čoraz častejšie používa širší termín - Pipeline (dopravník), ktorý opisuje akýkoľvek sekvenčný reťazec spracovania, kde sa výstup jednej etapy stáva vstupom pre ďalšiu.

Tento prístup sa vzťahuje nielen na údaje, ale aj na iné typy automatizácie: spracovanie úloh, vytváranie správ, integrácia so softvérom a digitálny pracovný tok (obr. 7.3-1).



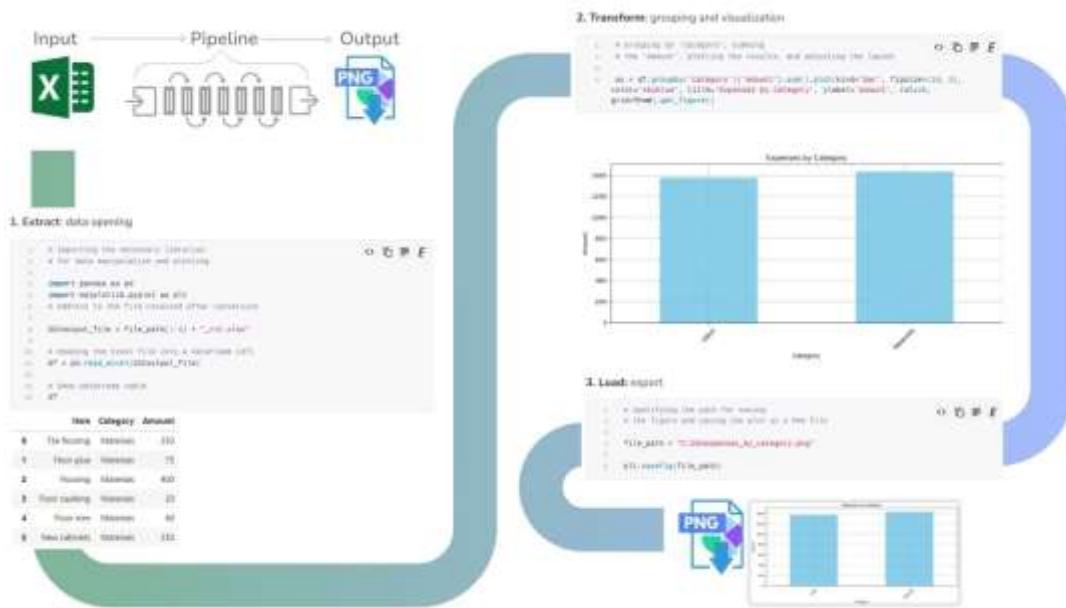
Obrázok 7.3-1 Pipeline je postupnosť spracovania, v ktorej sa výstup jedného stupňa stáva vstupom do ďalšieho stupňa.

Používanie pipeline je jedným z hlavných prvkov automatizácie, najmä pri práci s veľkým množstvom heterogénnych údajov. Architektúra Pipeline umožňuje organizovať zložité kroky spracovania v modulárnom, konzistentnom a spravovateľnom formáte, čo zvyšuje čitateľnosť, zjednodušuje údržbu kódu a umožňuje inkrementálne ladenie a škálovateľné testovanie.



Obr. 7.3-2 ROI Proces validácie údajov v potrubnej sieti skracuje čas vykonávania o desiatky a stovky krát v porovnaní so spracovaním pomocou klasických nástrojov [74].

Na rozdiel od manuálnej práce v proprietárnych systémoch (ERP, PMIS, CAD, atď.), pipelining umožňuje výrazne (obr. 7.3-2) zvýšiť rýchlosť úloh, vyhnúť sa opakovanej práci a automatizovať spustenie procesov v správnom čase (obr. 7.3-3).

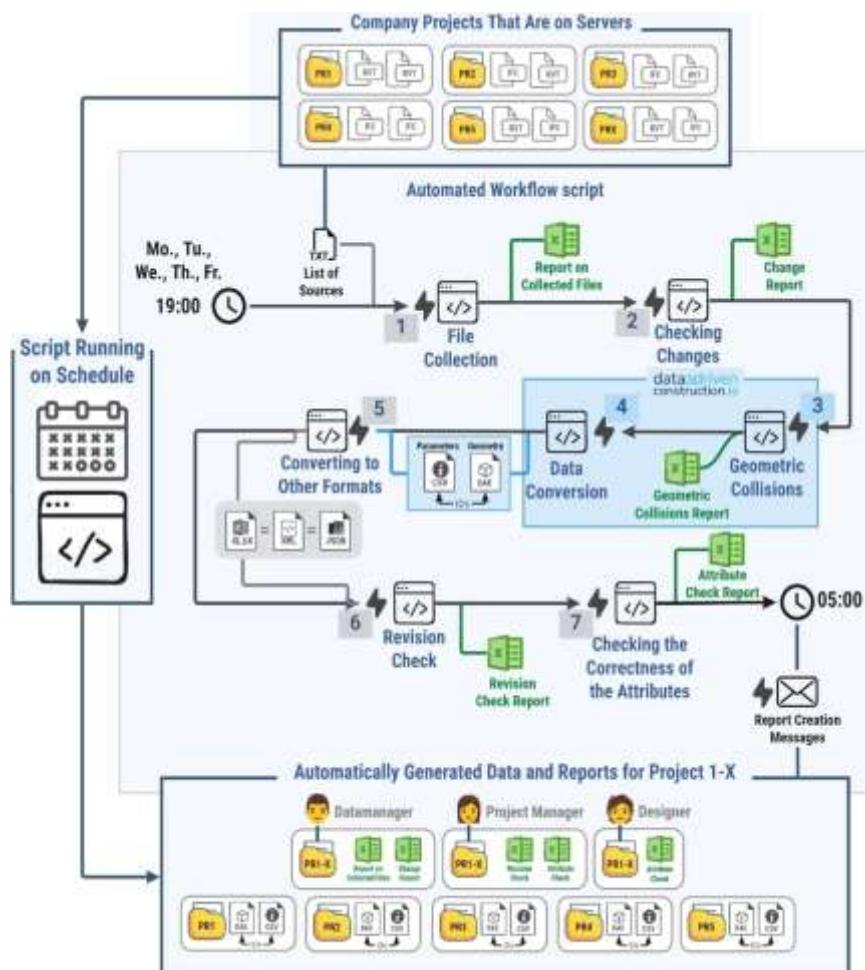


Obr. 7.3-3 Príklad ETL Potrubie na automatické odvodenie grafu z tabuľkových údajov v súbore XLSX bez otvorenia programu Excel.

Ak chcete spracovať prúdové údaje a vytvoriť automatizovanú Potrubie, podobne ako pri procese ETL, musíte vopred určiť zdroje údajov, ako aj časový rámec ich zberu - buď pre konkrétny obchodný proces, alebo pre celú spoločnosť.

V stavebných projektoch pochádzajú údaje z mnohých rôznych zdrojov s rôznymi intervalmi aktualizácie. Na vytvorenie spoľahlivej ukážky údajov je veľmi dôležité zaznamenávať, kedy sa informácie získavajú a aktualizujú. To umožňuje včasné rozhodovanie a zvyšuje efektívnosť riadenia projektu.

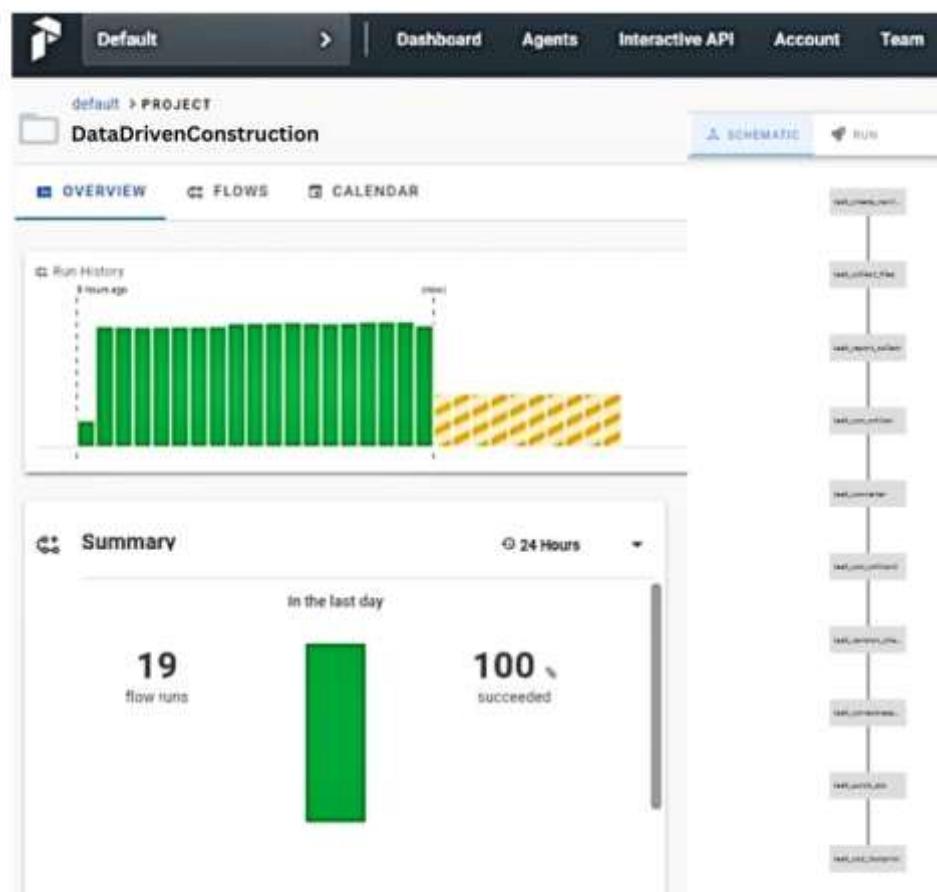
Jednou z možností je začať proces montáže v pevne stanovenom čase - napríklad o 19:00 na konci pracovného dňa. V tomto okamihu sa aktivuje prvý skript zodpovedný za agregáciu údajov z rôznych systémov a úložísk (obr. 7.3-4 krok 1). Potom nasleduje automatické spracovanie a transformácia údajov do štruktúrovaného formátu vhodného na analýzu (obr. 7.3-4 krok 2-4). V záverečnej fáze sa pomocou pripravených údajov automaticky generujú reporty, dashboardy a ďalšie produkty opísané v predchádzajúcich kapitolách (obr. 7.3-4 krok 6-7). Výsledkom je, že už o 5:00 ráno majú manažéri k dispozícii aktuálne správy o stave projektu v požadovanom formáte (obr. 7.3-5).



Obr. 7.3-4 Údaje v Pipeline, ktoré sa automaticky zbierajú večer, sa spracúvajú cez noc, takže manažéri majú aktuálne správy a ráno čerstvé hlásenia.

Včasný zber údajov, definovanie kľúčových ukazovateľov výkonnosti, automatizácia transformačných procesov a vizualizácia prostredníctvom informačných panelov sú kľúčovými prvkami úspešného rozhodovania založeného na údajoch.

Takéto automatizované procesy (obr. 7.3-4) sa môžu vykonávať úplne samostatne: bežia podľa plánu, spracúvajú údaje bez dozoru a môžu byť nasadené buď v cloude, alebo na vlastnom serveri spoločnosti (obr. 7.3-5). To umožňuje integrovať takéto potrubia ETL do existujúcej infraštruktúry IT, zachovať kontrolu nad údajmi a zabezpečiť flexibilitu pri škálovaní.



Obr. 7.3-5 Automatické procesy ETL -conveyor (obr. 7.3-4) na platforme Prefect, v ktorej sa každý pracovný deň po 19:00 striedavo spúšta 10 python skriptov.

Automatizácia pracovných postupov nielenže zvyšuje produktivitu tímu tým, že uvoľňuje čas na zmysluplnnejšie a menej rutinné úlohy, ale slúži aj ako dôležitý prvý krok k začleneniu technológií umelej inteligencie (AI) do podnikových procesov, ktorým sa budeme podrobnejšie venovať v kapitole o prediktívnej analýze a strojovom učení.

## Potrubie -ETL proces overovania údajov pomocou LLM

V predchádzajúcich kapitolách o vytváraní požiadaviek na údaje a automatizácii ETL, sme krok za krokom rozobrali proces prípravy, transformácie, validácie a vizualizácie údajov. Tieto činnosti boli implementované ako samostatné bloky kódu (obrázok 7.2-18 - obrázok 7.2-20), z ktorých každý vykonáva konkrétnu úlohu.

Teraz máme ďalší cieľ - spojiť tieto prvky do jednej ucelenej a automatizovanej pipeline spracovania údajov - pipeline, ETL -Pipeline - v ktorej sa všetky fázy (načítanie, validácia, vizualizácia, export) vykonávajú postupne v jednom automaticky spustiteľnom skripte.

V nasledujúcom príklade bude implementovaný kompletný cyklus spracovania údajov: od načítania zdrojového súboru CSV → až po kontrolu štruktúry a hodnôt pomocou regulárnych výrazov → výpočet výsledkov → generovanie vizuálnej správy vo formáte PDF.

💡 Na získanie príslušného kódu môžete použiť nasledujúci textový dotaz do LLM:

Napíšte prosím vzorku kódu, ktorá načíta údaje z CSV, overí DataFrame pomocou regulárnych výrazov, skontroluje identifikátory vo formáte "W-NEW" alebo "W-OLD", energetickú účinnosť s písmenami "A" až "G", záručnú dobu a cyklus výmeny s číselnými hodnotami v rokoch a na konci vytvorí správu s počtom vyhovujúcich a nevyhovujúcich hodnôt, vygeneruje PDF s histogramom výsledkov a pridá textový popis. ↵

💡 Odpoved' LLM :

```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}\$', 
14     'Energy_Efficiency': r'^[A-G]\+\$\$', 
15     'Warranty_Period': r'^\d+\syears\$', 
16     'Replacement_Cycle': r'^\d+\syears\$'
17 }
18
19 for column, pattern in patterns.items():
20     df['verified_' + column] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Obr. 7.3-6 Pipeline (ETL) automatizuje celý cyklus spracovania údajov: od stiahnutia a validácie až po vytvorenie štruktúrovanej správy vo formáte PDF .

Automatizovaný kód (obr. 7.3-6) v chate LLM alebo v DIE po skopírovaní kódu overí údaje zo súboru CSV - pomocou zadaných regulárnych výrazov, vytvorí správu o počte vyhovujúcich a nevyhovujúcich záznamov a potom uloží výsledky overenia ako súbor PDF -.

Táto štruktúra ETL, kde je každý krok - od načítania údajov až po generovanie zostáv - implementovaný ako samostatný modul, zabezpečuje transparentnosť, škálovateľnosť a reprodukovateľnosť. Prezentácia logiky overovania ako ľahko čitateľného kódu v jazyku Python robí proces transparentným a zrozumiteľným nielen pre vývojárov, ale aj pre odborníkov na správu údajov, kvalitu a analýzu.

Prístup spoločnosti Pipeline k automatizácii spracovania údajov umožňuje štandardizovať procesy, zvýšiť ich opakovateľnosť a zjednodušíť prispôsobenie novým projektom. Vytvára sa tak jednotná metodika analýzy údajov bez ohľadu na zdroj alebo typ úlohy - či už ide o testovanie zhody, generovanie správ alebo prenos údajov do externých systémov.

Takáto automatizácia znižuje chybovosť ľudského faktora, znižuje závislosť od vlastných riešení a zvyšuje presnosť a spoľahlivosť výsledkov, vďaka čomu sú vhodné na operatívnu analýzu na úrovni projektu aj na strategickú analýzu na úrovni spoločnosti.

## **Potrubie -ETL: overenie údajov a informácií o prvkoch projektu v CAD (BIM)**

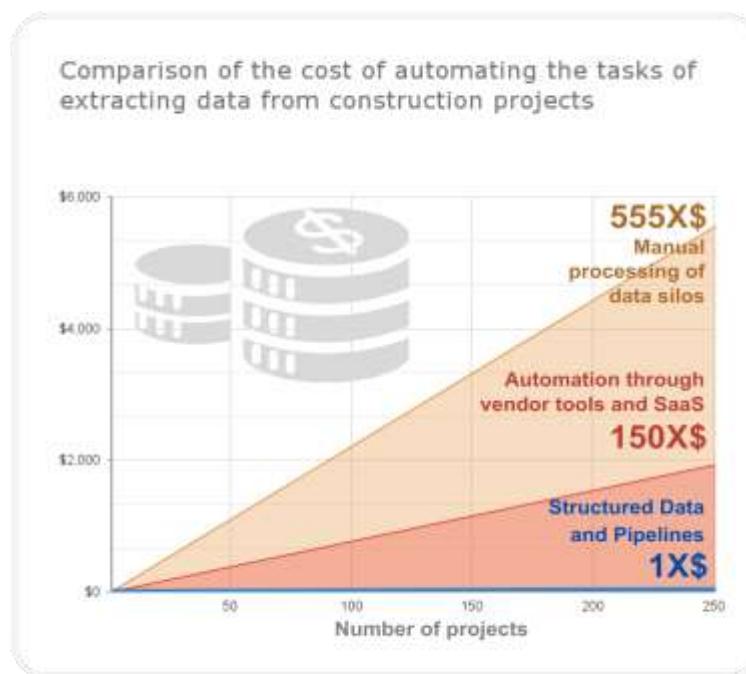
Údaje zo systémov CAD a databáz (BIM) patria medzi naj sofistikovanejšie a najdynamickejšie aktualizované zdroje údajov v podnikaní stavebných spoločností. Tieto aplikácie nielenže opisujú projekt pomocou geometrie, ale ho aj dopĺňajú o viaceré vrstvy textových informácií: objemy, vlastnosti materiálov, priradenie miestností, úrovne energetickej účinnosti, tolerancie, predpokladanú životnosť a ďalšie atribúty.

Atribúty priradené entitám v CAD -modeloch sa vytvárajú vo fáze návrhu a stávajú sa základom pre ďalšie obchodné procesy vrátane kalkulácie nákladov, plánovania, hodnotenia životného cyklu a integrácie so systémami ERP a CAFM -systémami, kde efektívnosť procesov do veľkej miery závisí od kvality údajov pochádzajúcich z oddelení návrhu.

Tradičný prístup k validácii atribútov v modeloch CAD- (BIM-) zahŕňa manuálnu validáciu (obr. 7.2-1), ktorá sa pri veľkom objeme modelov stáva zdĺhavým a nákladným procesom. Vzhľadom na objem a počet moderných stavebných projektov a ich pravidelnú aktualizáciu sa proces validácie a transformácie údajov stáva neudržateľným a cenovo nedostupným.

Generálni dodávatelia a projektoví manažéri musia spracúvať veľké množstvo údajov o projektoch vrátane viacerých verzií a fragmentov tých istých modelov. Údaje pochádzajú od projekčných organizácií vo formátoch RVT, DWG, DGN, IFC, NWD a iných (obrázok 3.1-14) a vyžadujú si pravidelnú kontrolu zhody s priemyselnými a podnikovými normami

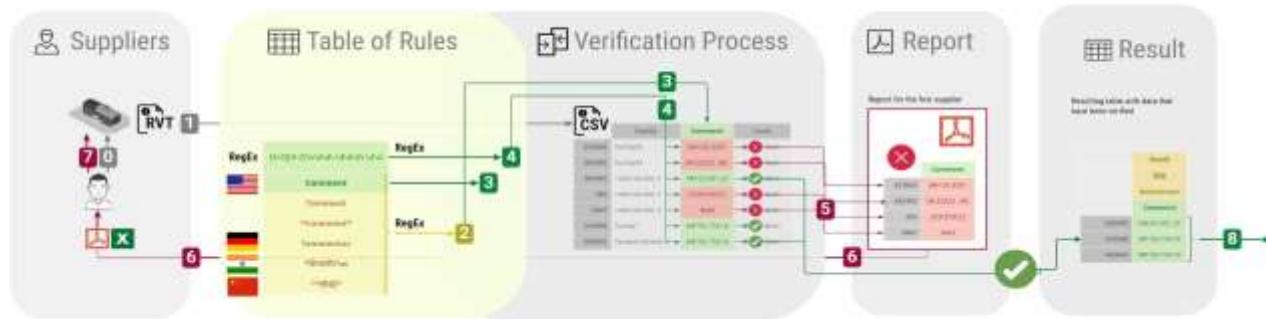
Závislosť od manuálnych činností a špecializovaného softvéru spôsobuje, že proces validácie údajov je úzkym miestom v pracovných postupoch týkajúcich sa údajov z celopodnikových modelov. Automatizácia a používanie štruktúrovaných požiadaviek môžu túto závislosť odstrániť, čím sa výrazne zvýši rýchlosť a spoľahlivosť validácie údajov (obrázok 7.3-7).



Obr. 7.3-7 Automatizácia zvyšuje rýchlosť overovania a spracovania údajov, čo znižuje náklady na prácu desiatky krát [140].

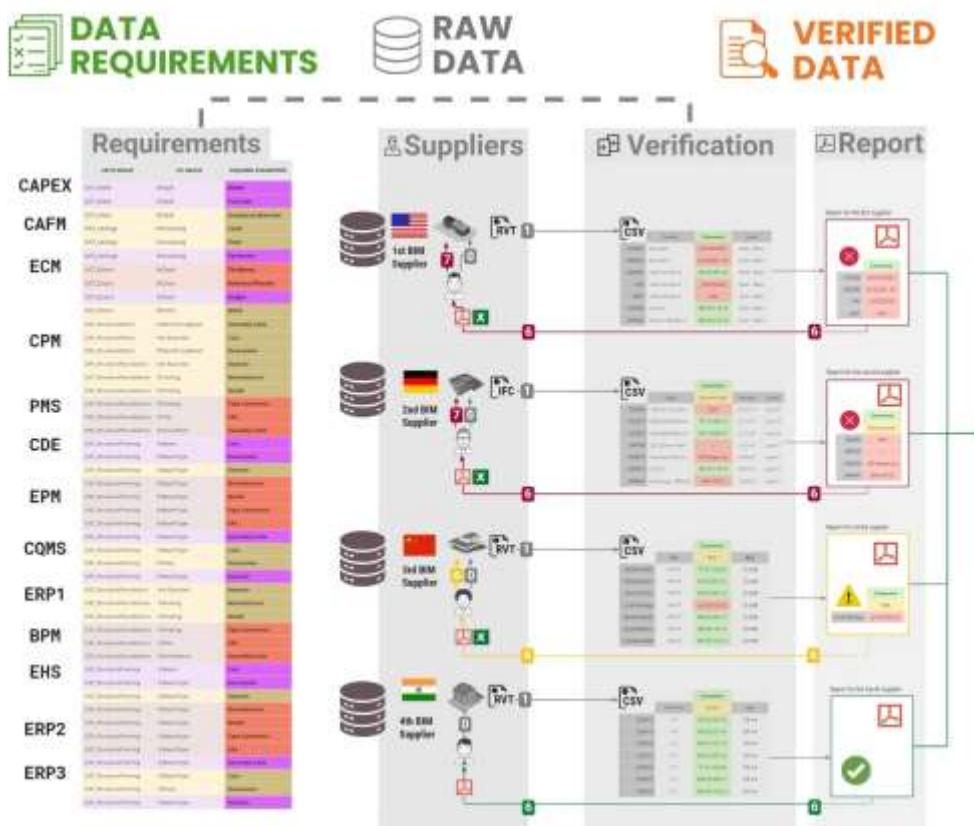
Proces validácie údajov CAD zahŕňa extrakciu údajov (fáza ETL Extract) z rôznych uzavretých (RVT, DWG, DGN, NWS atď.) alebo otvorených pološtruktúrovaných a parametrických formátov (IFC, CPXML, USD.) alebo otvorených pološtruktúrovaných a parametrických formátov (IFC, CPXML, USD), v ktorých možno na každý atribút a jeho hodnoty aplikovať tabuľky pravidiel (fáza Transformácia) pomocou regulárnych výrazov RegEx (obr. 7.3-8), pričom tomuto procesu sme sa podrobne venovali vo štvrtej časti knihy.

Vytvorenie chybového hlásenia vo formáte PDF o a úspešne validovaných záznamoch by malo byť ukončené výstupom (krok Load) v štruktúrovaných formátoch, ktoré zohľadňujú len validované entity, ktoré možno použiť na ďalšie procesy.



Obrázok 7.3-8 Proces validácie údajov od poskytovateľov údajov o projekte po záverečné správou validovanú pomocou regulárnych výrazov.

Automatizácia validácie údajov zo systémov CAD (BIM) so štruktúrovanými požiadavkami a streamovanie nových údajov, ktoré sa spracúvajú prostredníctvom ETL-Pipelines (obrázok 7.3-9), znížuje potrebu manuálnej účasti na procese validácie (jednotlivé procesy validácie a požiadaviek na údaje boli opísané v predchádzajúcich kapitolách).



Obr. 7.3-9 Automatizácia validácie údajov prostredníctvom ETL zjednodušuje riadenie stavebných projektov tým, že urýchľuje procesy.

Tradične môže overenie modelov poskytnutých dodávateľmi a špecialistami CAD (BIM) trvať dni až

týždne. Zavedením automatizovaných procesov ETL sa však tento čas môže skrátiť na niekoľko minút. V typickej situácii zhotoviteľ uvádza: "Model je validovaný a vyhovujúci." Týmto vyhlásením sa začína reťaz overovania tvrdenia zhotoviteľa o kvalite údajov:

- ☛ Projektový manažér - "Zhotoviteľ tvrdí: 'Model bol otestovaný, všetko je v poriadku.'"
- ☛ Správca údajov - Overenie načítania:
  - Jednoduchý skript v programe Pandas zistí porušenie v priebehu niekoľkých sekúnd. Automatizácia eliminuje spory:
    - Kategória: OST\_StructuralColumns, Parameter: FireRating IS NULL.
    - Generovanie zoznamu ID porušení → export do Excelu/PDF.

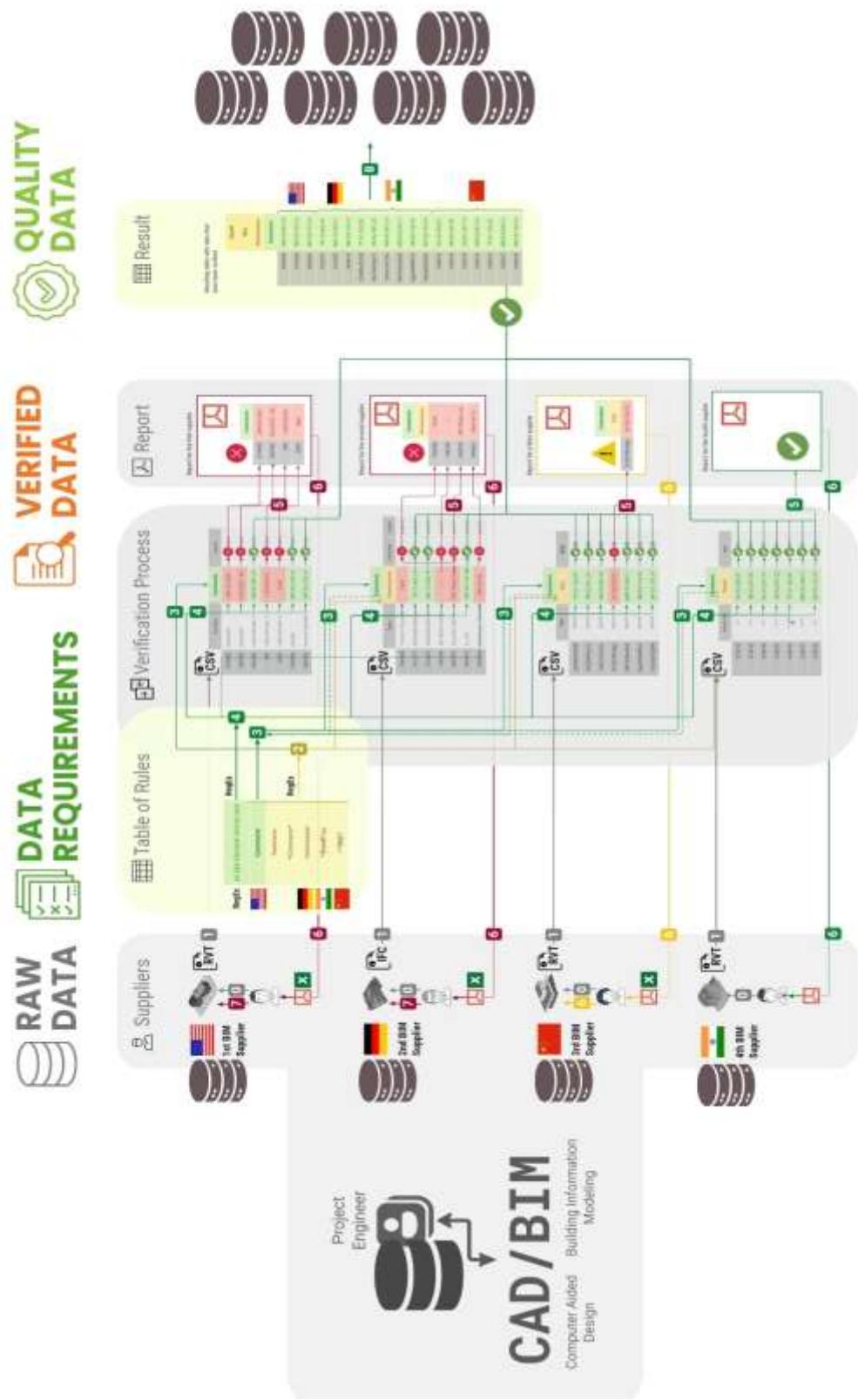
Jednoduchý skript v programe Pandas zistí narušenie v priebehu niekoľkých sekúnd:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Filtrovanie
issues = df[df["FireRating"].isnull()] # Prázdne hodnoty
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Export ID
```

- ☛ Manažér údajov projektovému manažérovi - "Kontrola stránky ukázala, že 18 stĺpcov nemá vyplnený parameter FireRating."
- ☛ Vedúci projektu zhotoviteľovi - "Model sa vracia na prepracovanie: parameter FireRating je povinný, bez neho nie je možné ho prevziať"

Výsledkom je, že model CAD neprechádza validáciou, automatizácia eliminuje spory a dodávateľ takmer okamžite dostane štruktúrovanú správu so zoznamom ID problematických prvkov. Týmto spôsobom sa proces validácie stáva transparentným, opakovateľným a chráneným pred ľudskou chybou (obr. 7.3-10).

Tento prístup mení proces validácie údajov na inžiniersku funkciu namiesto manuálnej kontroly kvality. Nielenže sa tým zvyšuje produktivita, ale umožňuje to aj aplikovať rovnakú logiku na všetky projekty spoločnosti, čo umožňuje komplexnú digitálnu transformáciu procesov od návrhu až po prevádzku.



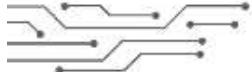
Obr. 7.3-10 Automatizácia kontroly atribútov prvkov eliminuje ľudské chyby a znižuje pravdepodobnosť chýb.

Prostredníctvom automatizovaných potrubí (obr. 7.3-10) môžu používatelia systému, ktorí očakávajú kvalitné údaje zo systémov CAD- (BIM-), okamžite získať výstupné údaje, ktoré potrebujú - tabuľky, dokumenty, obrázky - a rýchlo ich integrovať do svojich pracovných úloh.

Automatizácia riadenia, spracovania a analýzy je hnacou silou zmeny v prístupe k riadeniu stavebných projektov, najmä v interoperabilite rôznych systémov bez používania zložitých a drahých modulárnych proprietárnych systémov alebo riešení uzavretých dodávateľov.

Zatiaľ čo pojmy a marketingové skratky prichádzajú a odchádzajú, samotné procesy overovania požiadaviek na údaje zostanú navždy neoddeliteľnou súčasťou obchodných procesov. Namiesto vytvárania ďalších a ďalších špecializovaných formátov a štandardov by sa stavebníctvo malo radšej obrátiť na nástroje, ktoré už preukázali svoju účinnosť v iných odvetviach. V súčasnosti existujú výkonné platformy na automatizáciu spracovania údajov a integráciu procesov, ktoré umožňujú spoločnostiam výrazne skrátiť čas strávený rutinnými operáciami a minimalizovať chyby pri extrakcii, transformácii a načítaní.

Jedným z populárnych príkladov riešení na automatizáciu a orchestráciu procesov ETL je Apache Airflow, ktorý umožňuje organizovať zložité výpočtové procesy a spravovať potrubia ETL. Popri Airflow sa aktívne používajú aj ďalšie podobné riešenia, ako napríklad Apache NiFi na smerovanie a streamovanie údajov a n8n na automatizáciu obchodných procesov.



## KAPITOLA 7.4.

# ORCHESTRÁCIA ETL A PRACOVNÝCH POSTUPOV: PRAKTICKÉ RIEŠENIA

### DAG a Apache Airflow: automatizácia a orchestrácia pracovných postupov

Apache Airflow je bezplatná platforma s otvoreným zdrojovým kódom, ktorá je určená na automatizáciu, orchestráciu a monitorovanie pracovných tokov (ETL -conveyors).

Práca s veľkým množstvom údajov je potrebná každý deň:

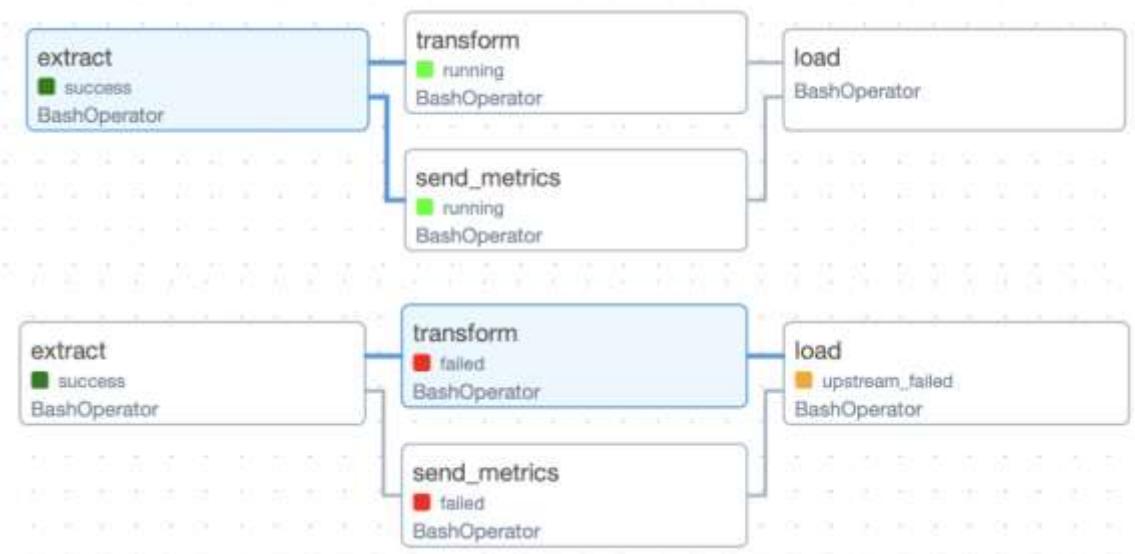
- Preberanie súborov z rôznych zdrojov - Výpis (napríklad od dodávateľov alebo zákazníkov).
- Transformácia týchto údajov do požadovaného formátu - Transformácia (štruktúra, čistenie a overenie)
- Odosielanie výsledkov na overenie a vytváranie správ - Načítanie (nahrávanie do požadovaných systémov, dokumentov, databáz alebo informačných panelov).

Manuálne vykonávanie takýchto procesov ETL zaberá veľa času a vedie k riziku ľudskej chyby. Zmena zdroja údajov alebo zlyhanie v niektorom z krokov môže spôsobiť oneskorenie a nesprávne výsledky.

Automatizačné nástroje, ako napríklad Apache Airflow, vám umožnia vytvoriť spoľahlivý ETL - prevodník, minimalizovať chyby, skrátiť čas spracovania a zabezpečiť správnosť údajov v každej fáze. Základom Apache Airflow je koncept DAG (Directed Acyclic Graph) - usmernený acyklický graf, v ktorom je každá úloha (operátor) spojená s inými závislosťami a vykonáva sa striktne v určenom poradí. DAG eliminuje cykly, čo poskytuje logickú a predvídateľnú štruktúru vykonávania úloh.

Airflow sa stará o orchestráciu - spravuje závislosti medzi úlohami, riadi harmonogramy vykonávania, sleduje stav a automaticky reaguje na zlyhania. Tento prístup minimalizuje manuálne zásahy a zabezpečuje spoľahlivosť celého procesu.

**Orchestrátor úloh** je nástroj alebo systém určený na správu a riadenie vykonávania úloh v komplexných počítačových a informačných prostrediah. Uľahčuje proces nasadzovania, automatizácie a riadenia vykonávania úloh s cieľom zvýšiť výkon a optimalizovať zdroje.



Obrázok 7.4-1 Apache Airflow poskytuje používateľsky prívetivé rozhranie, v ktorom môžete vizualizovať DAG -ETL, zobraziť protokoly vykonávania, stav spustenia úloh a ďalšie informácie.

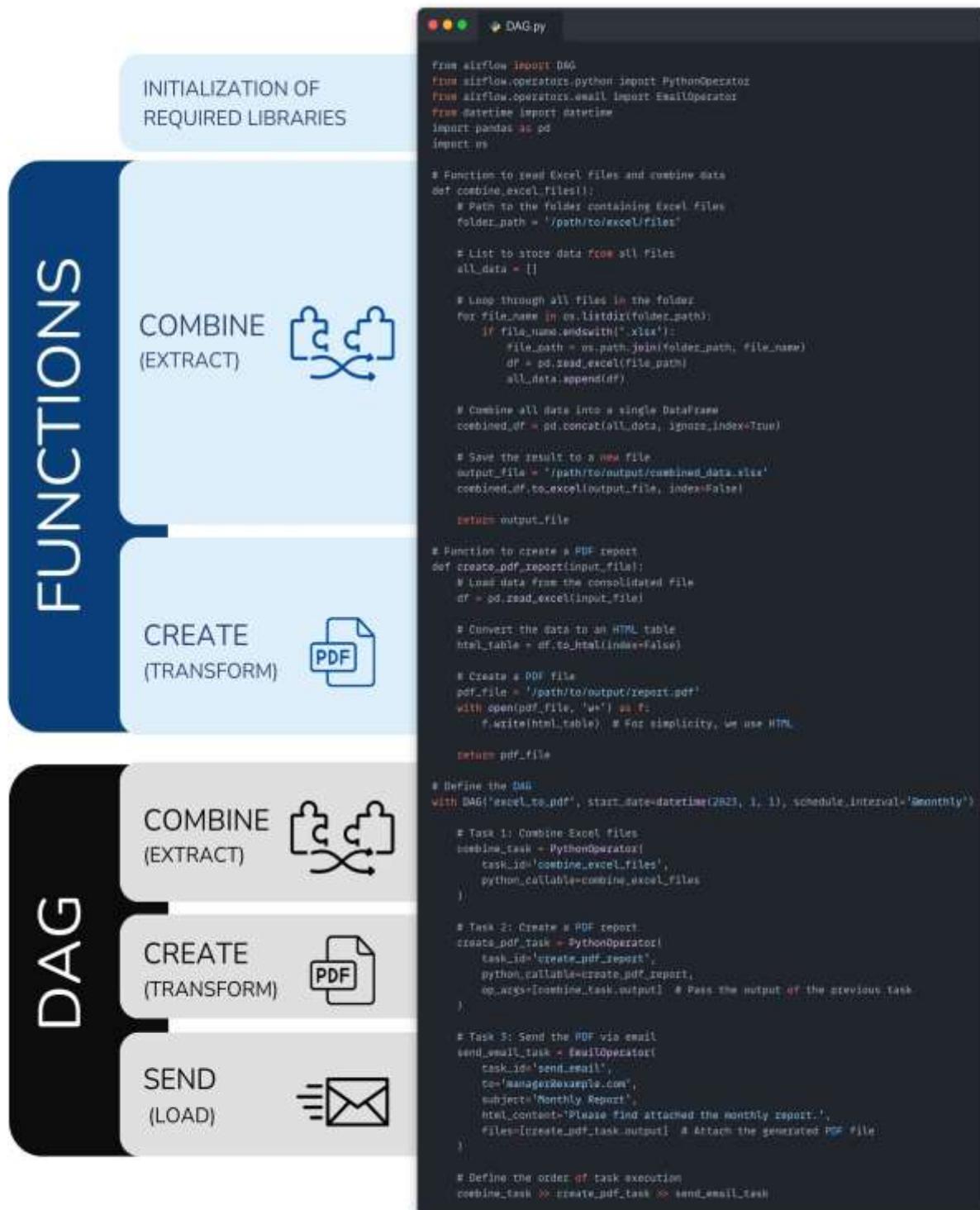
Airflow sa široko používa na orchestráciu a automatizáciu distribuovaných výpočtov, spracovanie údajov, riadenie procesov ETL (Extract, Transform, Load), plánovanie úloh a iné dátové scenáre. V predvolenom nastavení Apache Airflow používa ako databázu SQLite.

Príklad jednoduchej DAG, podobne ako ETL, pozostáva z úloh - Extract, Transform a Load. V grafe, ktorý sa ovláda prostredníctvom používateľského rozhrania (obr. 7.4-1), je definované poradie vykonávania úloh (fragmentov kódu): napríklad najprv sa vykoná extrakcia, potom transformácia (a odoslanie metriky) a prácu dokončí úloha load. Keď sa dokončia všetky úlohy, proces načítania údajov sa považuje za úspešný.

## Apache Airflow: praktická aplikácia automatizáciu ETL

Apache Airflow je široko používaný na organizáciu komplexných procesov spracovania údajov a umožňuje vytvárať flexibilné ETL -konvektory. Apache Airflow možno spustiť buď cez webové rozhranie, alebo programovo prostredníctvom kódu jazyka Python (obr. 7.4-2). Vo webovom rozhraní (obr. 7.4-3) môžu správcovia a vývojári vizuálne sledovať DAG, spúštať úlohy a analyzovať výsledky vykonávania.

Pomocou DAG, môžete nastaviť jasnú postupnosť úloh, spravovať závislosti medzi nimi a automaticky reagovať na zmeny zdrojových údajov. Uvedme príklad použitia Airflow na automatizáciu spracovania hlásení (obr. 7.4-2).



Obr. 7.4-2 Koncepcia ETL - dopravník na spracovanie údajov pomocou Apache Airflow.

V tomto príklade (obrázok 7.4-2) sa uvažuje o DAG, ktorá vykonáva kľúčové úlohy v rámci ETL -

ETL (EXTRACT, TRANSFORM, LOAD) PIPELINE

prevodníka:

■ **Čítanie súborov Excel (Výpis):**

- Sekvenčné prechádzanie všetkých súborov v danom adresári.
- Čítanie údajov z každého súboru pomocou knižnice pandas.
- Zlúčiť všetky údaje do jedného DataFrame.

■ **Vytvoriť PDF -dokument (Transformácia):**

- Transformácia zlúčeného DataFrame do HTML -tabuľky.
- Uložte tabuľku ako PDF (v demonštračnej verzii - prostredníctvom HTML).

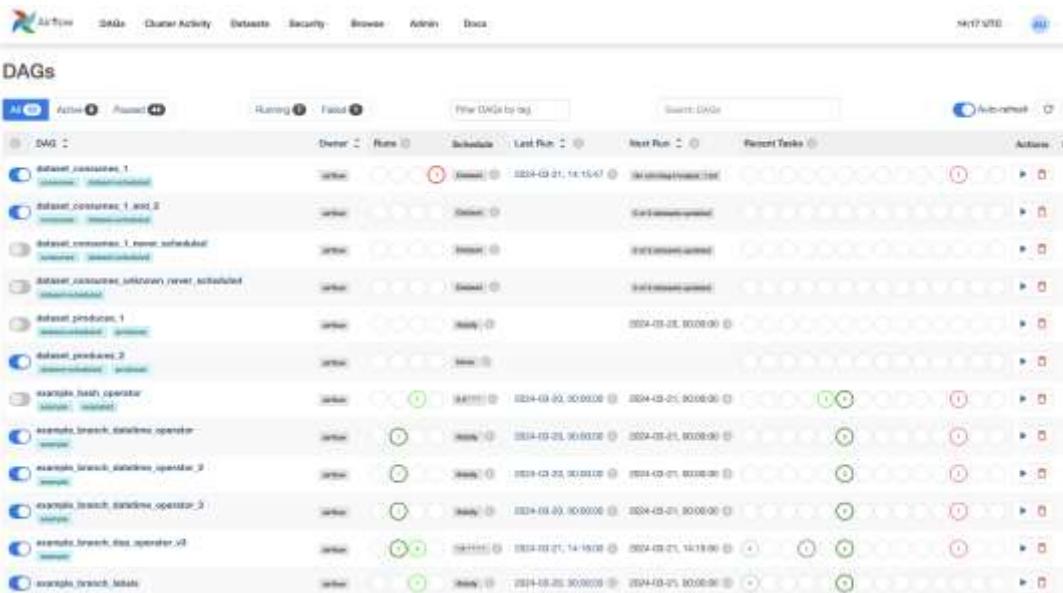
■ **Odoslanie správy e-mailom (Load):**

- Použite EmailOperator na odoslanie PDF -dokumentu e-mailom.

■ **Konfigurácia DAG:**

- Definovanie postupnosti úloh: extrakcia údajov → generovanie správy → odosielanie.
- Priradenie plánu spustenia (@mesačne - prvý deň každého mesiaca).

Automatický príklad ETL (obrázok 7.4-2) ukazuje, ako zbierať údaje zo súborov Excel, vytvoriť dokument PDF a poslať ho e-mailom. Toto je len jeden z mnohých možných scenárov použitia Airflow. Tento príklad je možné prispôsobiť akejkoľvek konkrétnej úlohe s cieľom zjednodušiť a automatizovať spracovanie údajov.



Obr. 7.4-3 Prehľad všetkých DAG v prostredí s informáciami o posledných spusteniach.

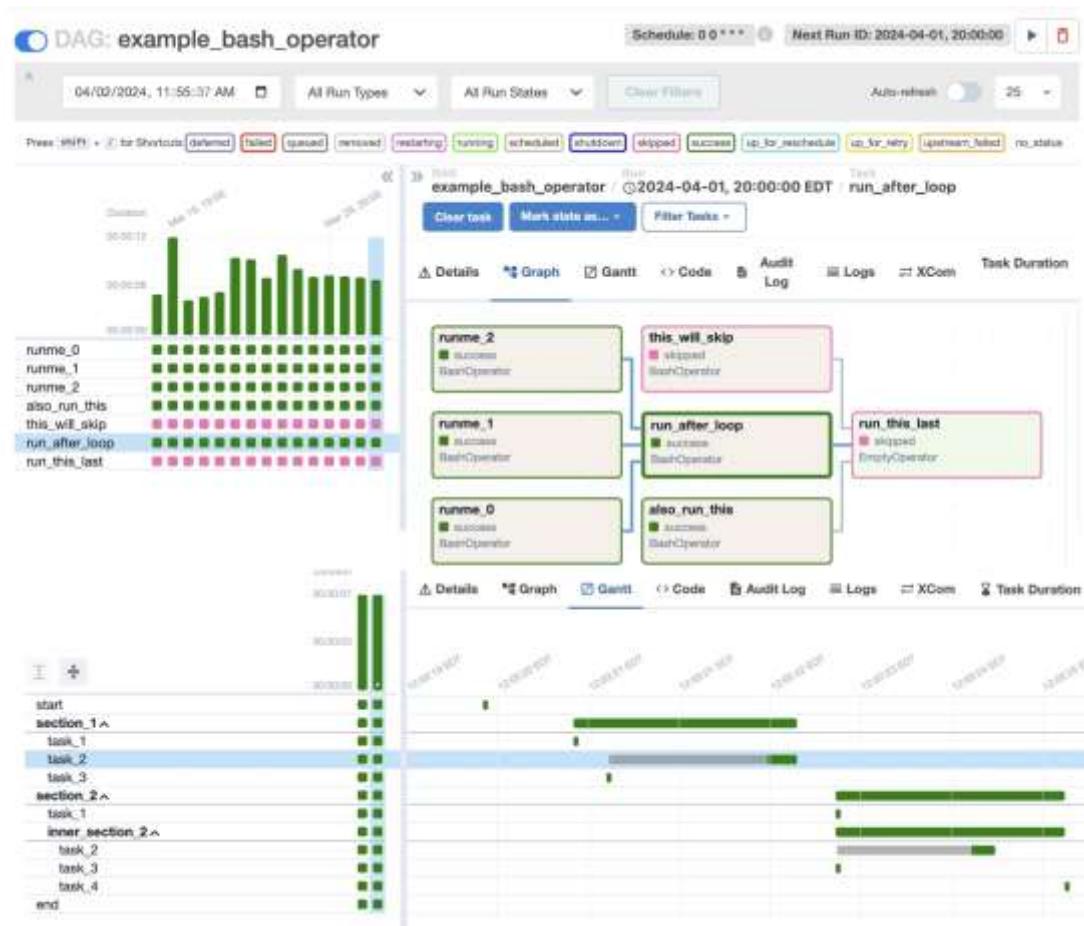
Webové rozhranie Apache Airflow (obrázok 7.4-3) poskytuje komplexné vizuálne prostredie na správu pracovných tokov údajov. Zobrazuje DAG ako interaktívne grafy, kde uzly predstavujú úlohy a hrany závislosti medzi nimi, čo uľahčuje sledovanie zložitých pracovných tokov údajov. Rozhranie obsahuje ovládací panel s informáciami o stave vykonávania úloh, histórii behu, podrobných protokoloch a výkonnostných metrikách. Správcovia môžu manuálne spúštať úlohy, reštartovať neúspešné operácie, pozastaviť DAG a konfigurovať premenné prostredia, a to všetko

prostredníctvom intuitívneho používateľského rozhrania.

Takáto architektúra môže byť doplnená o validáciu údajov, oznámenia o stave vykonávania, integráciu s externými API alebo databázami. Airflow umožňuje flexibilne prispôsobovať DAG: pridávať nové úlohy, meniť ich poradie, kombinovať reťazce - čo z neho robí efektívny nástroj na automatizáciu komplexných procesov spracovania údajov. Pri spustení DAG vo webovom rozhraní Airflow (obr. 7.4-3, obr. 7.4-4) môžete sledovať stav úloh. Systém používa farebnú indikáciu:

- Zelená - úloha bola úspešne dokončená.
- Žltá - proces prebieha.
- Červená - chyba pri vykonávaní úlohy.

V prípade zlyhania (napr. chýbajúci súbor alebo poškodená dátová štruktúra) systém automaticky iniciuje odoslanie oznámenia.



Obrázok 7.4-4 Apache Airflow výrazne zjednodušuje diagnostiku problémov, optimalizáciu procesov a tímovú spoluprácu na zložitých potrubiac spracovania dát.

Apache Airflow je pohodlný, pretože automatizuje rutinné úlohy a eliminuje potrebu ich manuálneho vykonávania. Zabezpečuje spoľahlivosť monitorovaním vykonávania procesov a okamžitým oznamovaním chýb. Flexibilita systému umožňuje ľahko pridať nové úlohy alebo upravovať existujúce, čím sa pracovné postupy prispôsobujú meniacim sa požiadavkám.

Okrem Apache Airflow, existujú podobné nástroje na orchestráciu pracovných tokov. Napríklad open source a bezplatný Prefect (obrázok 7.3-5) ponúka jednoduchšiu syntax a lepšie sa integruje s Pythonom, Luigi, vyvinutý spoločnosťou Spotify, poskytuje podobnú funkcionality a dobre pracuje s veľkými dátami. Za zmienku stojí aj Kronos a Dagster, ktoré ponúkajú moderné prístupy k budovaniu Pipeline so zameraním na modularitu a škálovateľnosť. Výber nástroja na orchestráciu úloh závisí od konkrétnych potrieb projektu, ale všetky pomáhajú automatizovať zložité dátové procesy ETL.

Osobitnú pozornosť si zaslúži Apache NiFi, platforma s otvoreným zdrojovým kódom, ktorá je určená na streamovanie a smerovanie údajov. Na rozdiel od Airflow, ktorý sa zameriava na dávkové spracovanie a správu závislostí, NiFi sa zameriava na transformáciu údajov v reálnom čase, za chodu a flexibilné smerovanie medzi systémami.

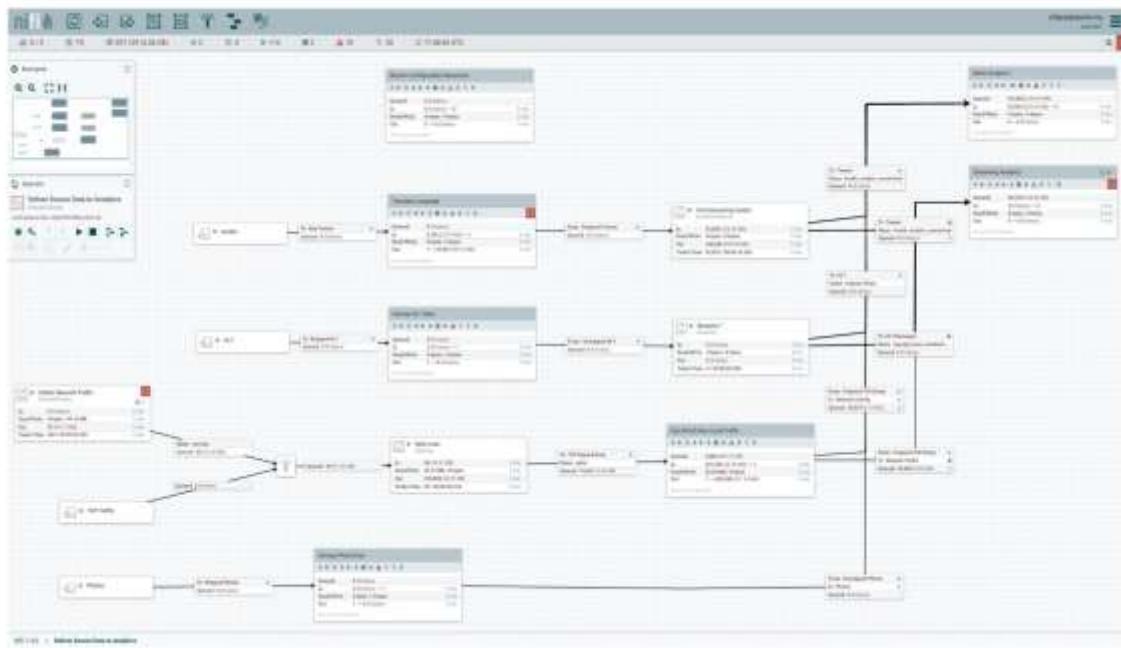
## Apache NiFi na smerovanie a konverziu údajov

Apache NiFi je výkonná open source platforma, určená na automatizáciu dátových tokov medzi rôznymi systémami. Pôvodne ju v roku 2006 vyvinula americká Národná bezpečnostná agentúra (NSA) pod názvom "Niagara Files" na interné použitie. V roku 2014 bol projekt open-sourcovaný a prevedený na Apache Software Foundation, čím sa stal súčasťou ich iniciatív v oblasti transferu technológií [141].

Apache NiFi je určený na zber, spracovanie a prenos údajov v reálnom čase. Na rozdiel od Airflow, ktorý pracuje s dávkovými úlohami a vyžaduje presne definované plány, NiFi pracuje v režime prúdového spracovania, čo umožňuje nepretržitý prenos údajov medzi rôznymi službami.

Apache NiFi je ideálny na integráciu so zariadeniami IoT, stavebnými senzormi, monitorovacími systémami a napr. na streamované overovanie formátov CAD na serveri, kde sa môže vyžadovať okamžitá reakcia na zmeny údajov.

Vďaka zabudovaným nástrojom na filtrovanie, transformáciu a smerovanie umožňuje NiFi štandardizáciu údajov (Transform) pred ich prenosom (Load) do úložných alebo analytických systémov. Jednou z jeho hlavných výhod je zabudovaná podpora zabezpečenia a kontroly prístupu, vďaka čomu je spoľahlivým riešením na spracovanie citlivých informácií.



Obrázok 7.4-5 Grafické znázornenie toku údajov v rozhraní Apache NiFi.

Apache NiFi efektívne spracováva úlohy streamovania, filtrovania a smerovania dát v reálnom čase. Je ideálny pre technicky náročné scenáre, kde je dôležitý stabilný prenos informácií medzi systémami a vysoká prieplustnosť.

Ak je však hlavným cieľom integrácia rôznych služieb, automatizácia rutinných operácií a rýchle nastavenie pracovných postupov bez hlbokých znalostí programovania, sú žiadane riešenia s nízkym vstupným prahom a maximálnou flexibilitou. Jedným z takýchto nástrojov je n8n - platforma triedy Low-Code /No-Code zameraná na automatizáciu biznisu a vizuálnu orchestráciu procesov.

## n8n Low-Code, No-Code orchestrácia procesov

**n8n** je open source platforma s nízkym kódom / No-Code na vytváranie automatizovaných pracovných postupov, ktorá sa vyznačuje jednoduchým používaním, flexibilitou a schopnosťou rýchlej integrácie so širokou škálou externých služieb.

**No-Code** je metóda vytvárania digitálnych produktov bez písania kódu. Všetky prvky procesu - od logiky až po rozhranie - sa realizujú výlučne pomocou vizuálnych nástrojov. No-Code platformy sú určené používateľom bez technického zázemia a umožňujú rýchlo vytvárať automatizácie, formuláre, integrácie a webové aplikácie. Príklad: používateľ nastaví automatické odosielanie oznámení alebo integráciu s tabuľkami Google prostredníctvom rozhrania drag-and-drop bez znalosti programovania.

Vďaka otvorenému zdrojovému kódu a možnostiam lokálneho nasadenia poskytuje n8n v procesoch

automatizácie a vytvárania potrubí ETL firmám úplnú kontrolu nad ich údajmi a zároveň zabezpečuje bezpečnosť a nezávislosť od poskytovateľov clodu.

Na rozdiel od Apache Airflow, ktorý je orientovaný na výpočtové úlohy s pevnou orkestráciou a vyžaduje si znalosť jazyka Python, n8n poskytuje vizuálny editor, ktorý umožňuje skriptovanie bez potreby znalosti programovacích jazykov (obrázok 7.4-6). Hoci jeho rozhranie umožňuje vytvárať automatizované procesy bez písania kódu (No-Code), v zložitejších scenároch môžu používateľia pridávať vlastné JavaScript a Python -funkcie na rozšírenie možností (Low-Code).

**Low-Code** je prístup k vývoju softvéru, pri ktorom sa základná logika aplikácie alebo procesu vytvára pomocou grafického rozhrania a vizuálnych prvkov a programový kód sa používa len na prispôsobenie alebo rozšírenie funkcií. Platformy Low-Code umožňujú výrazne urýchliť vývoj riešení tým, že sa do nich zapájajú nielen programátori, ale aj podnikoví používateľia so základnými technickými zručnosťami. Príklad: používateľ môže zostaviť obchodný proces z pripravených blokov a v prípade potreby pridať vlastný skript v jazyku JavaScript alebo Python.

Hoci je n8n umiestnený ako platforma s nízkym vstupným prahom, základné znalosti programovania, porozumenie webovým technológiám a zručnosti v práci s API. Flexibilita systému umožňuje jeho prispôsobenie na širokú škálu úloh - od automatizovaného spracovania údajov až po integráciu s messengermi, zariadeniami internetu vecí a cloudovými službami.

Hlavné funkcie a výhody používania zariadenia n8n:

- Možnosti nasadenia s **otvoreným zdrojovým kódom** a lokálne nasadenie zabezpečujú úplnú kontrolu nad údajmi, dodržiavanie bezpečnosti a nezávislosť od poskytovateľov cludových služieb.
- **Integrácia s viac ako 330 službami** vrátane CRM, ERP, elektronického obchodu, cloudových platoform, messengerov a databáz.
- **Flexibilita scenárov:** od jednoduchých oznámení až po komplexné reťazce so spracovaním požiadaviek API, logikou rozhodovania a pripojením služieb AI.
- **Podpora pre JavaScript a Python:** používateľia môžu podľa potreby vkladať vlastný kód a rozširovať tak možnosti automatizácie.
- **Intuitívne vizuálne rozhranie:** umožňuje rýchlo konfigurovať a vizualizovať všetky kroky procesu.

Platformy triedy Low-Code poskytujú nástroje na vytváranie digitálnych riešení s minimálnym množstvom kódu, takže sú ideálne pre tímy, ktoré nemajú hlboké technické znalosti, ale potrebujú automatizovať procesy.

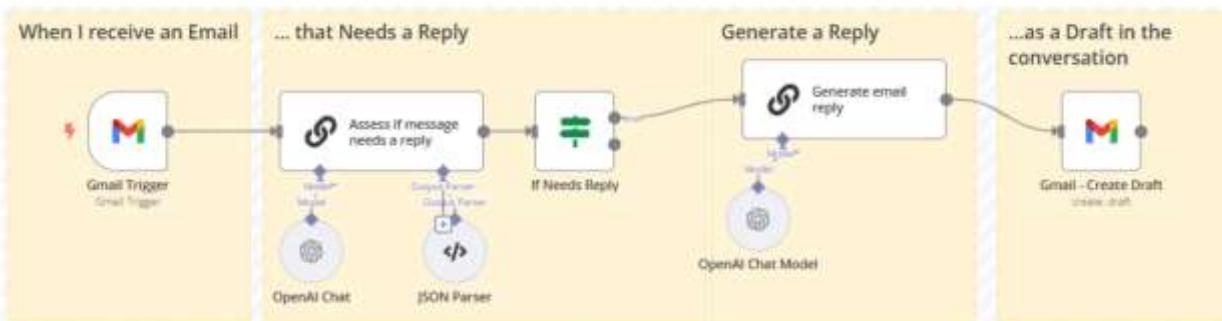
V stavebníctve možno n8n použiť na automatizáciu rôznych procesov, ako je integrácia so systémami riadenia projektov, kontrola toku, písanie hotových správ a listov, automatická aktualizácia údajov o zásobách materiálu, zasielanie oznámení o stave úloh tímom a ďalšie. Prispôsobená potrubná linka v systéme n8n môže niekoľkonásobne znížiť počet manuálnych operácií, znížiť pravdepodobnosť chýb

a urýchliť rozhodovanie o realizácii projektu.

Môžete si vybrať z takmer dvoch tisícok pripravených, bezplatných a otvorených zdrojov n8n Pipeline, ktoré sú k dispozícii na adrese: [n8n.io/workflows](https://n8n.io/workflows), na automatizáciu stavebných pracovných postupov aj osobných úloh, čím sa zníži počet rutinných operácií.

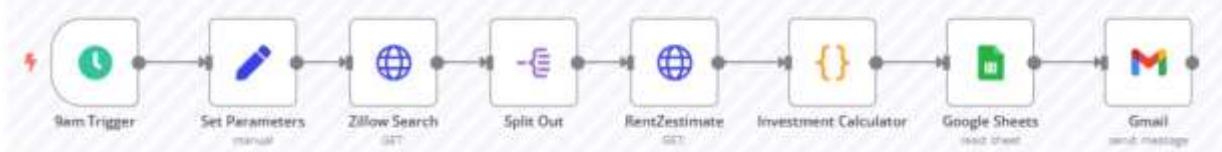
Vezmite si jednu z pripravených šablón Pipeline, ktorá je bezplatne k dispozícii na adrese [n8n.io](https://n8n.io) [142] a ktorá automaticky vytvára návrhy odpovedí v službe Gmail (obr. 7.4-6), čím pomáha používateľom, ktorí dostávajú veľké množstvo e-mailov alebo majú problémy s tvorbou odpovedí.

Táto šablóna n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Obrázok 7.4-6) analyzuje prichádzajúce e-maily pomocou LLM z ChatGPT, určí, či je potrebná odpoveď, vytvorí koncept z ChatGPT a prevedie text do HTML a pridá ho do reťazca správ v službe Gmail. Tým sa e-mail automaticky neodošle, čo umožňuje ručne upraviť a schváliť odpoveď. Nastavenie trvá približne 10 minút a zahŕňa konfiguráciu OAuth rozhrania API služby Gmail a integráciu rozhrania API OpenAI. Výsledkom je pohodlné a bezplatné riešenie na automatizáciu bežnej e-mailovej komunikácie bez straty kontroly nad obsahom e-mailov.



Obr. 7.4-6 Automatizovaný proces generovania e-mailovej odpovede pomocou n8n.

Ďalším príkladom automatizácie pomocou n8n je vyhľadávanie výhodných ponúk na trhu s nehnuteľnosťami [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", zhromažďuje denne relevantné ponuky zodpovedajúce zadaným kritériám pomocou rozhrania Zillow API. Automaticky vypočítava kľúčové investičné ukazovatele (Cash on Cash ROI, Monthly Cash Flow, Down Payment), aktualizuje Google Sheets a posiela súhrnnú správu na e-mail (obrázok 7.4-7), čo investorom umožňuje ušetriť čas a rýchlo reagovať na najlepšie ponuky.



Obr. 7.4-7 Automatizovaný proces posudzovania investičnej atraktívnosti nehnuteľností.

Vďaka svojej flexibilite a rozšíriteľnosti je n8n cenným nástrojom pre spoločnosti, ktoré sa snažia digitálne transformovať a stať sa konkurencieschopnejšími na trhu pomocou relatívne jednoduchých

a bezplatných nástrojov s otvoreným zdrojovým kódom.

Nástroje ako Apache NiFi, Airflow a n8n možno považovať za tri vrstvy spracovania údajov (obrázok 7.4-8). NiFi riadi tok údajov, zabezpečuje ich doručenie a transformáciu, Airflow organizuje vykonávanie úloh agregovaním údajov do spracovateľských potrubí a n8n automatizuje integráciu s externými službami a riadi obchodnú logiku.

	The main task	Approach
<b>Apache NiFi</b>	Streaming and data transformation	Real-time stream processing
<b>Apache Airflow</b>	Task orchestration, ETL pipelines	Batch planning, DAG processes
<b>n8n</b>	Integration, automation of business logic	Low-code visual orchestration

Obr. 7.4-8 Apache Airflow, Apache NiFi a n8n možno považovať za tri navzájom sa dopĺňajúce vrstvy modernej architektúry správy údajov.

Tieto bezplatné nástroje s otvoreným zdrojovým kódom spolu potenciálne tvoria príklad efektívneho ekosystému pre správu údajov a procesov v stavebníctve, ktorý umožňuje spoločnostiam využívať informácie na rozhodovanie a automatizáciu procesov.

## Ďalšie kroky: prechod od manuálnych operácií k riešeniam založeným na analytike

Dnešné stavebné spoločnosti pôsobia v prostredí vysokej neistoty: meniace sa ceny materiálov, oneskorené dodávky, nedostatok pracovných síl a krátke termíny projektov. Využívanie analytických informačných panelov, ETL -konvertorov a systémov BI pomáha spoločnostiam rýchlo identifikovať problémové oblasti, posúdiť efektívnosť zdrojov a predvídať zmeny skôr, ako povedú k finančným stratám.

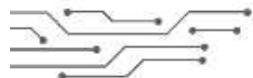
Na zhrnutie tejto časti stojí za to zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované technológie vo vašich každodenných úlohách:

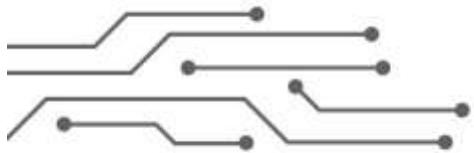
- Implementácia vizualizácií údajov a analytických panelov
  - Ovládanie procesu vytvárania informačných panelov na monitorovanie kľúčových ukazovateľov výkonnosti (KPI)
  - Používajte nástroje na vizualizáciu údajov (Power BI, Tableau, Matplotlib, Plotly).
- Automatizácia spracovania údajov prostredníctvom procesov ETL

- Nastavenie automatického zberu údajov z rôznych zdrojov (dokumentácia, tabuľky, CAD) prostredníctvom procesov ETL
  - Organizovať transformáciu údajov (napr. kontrolu regulárnych výrazov alebo výpočet) pomocou skriptov Python
  - Vyskúšajte nastavenie automatického vytvárania správ PDF (alebo DOC) pomocou knižnice FPDF s použitím údajov zo súborov Excel alebo extrahovaním informácií z iných dokumentov PDF.
- Používanie jazykových modelov (LLM) na automatizáciu
- používanie rozsiahlych jazykových modelov (LLM) na generovanie kódu, ktorý pomáha extrahovať a analyzovať údaje z neštruktúrovaných dokumentov
  - Zoznámte sa s nástrojom na automatizáciu n8n a preskúmajte pripravené šablóny a prípadové štúdie na ich webovej stránke. Určite, ktoré procesy z vašej práce možno plne automatizovať pomocou prístupu No-Code/Low-Code.

Analytický prístup k údajom a automatizácii procesov nielen skracuje čas strávený rutinnými operáciami, ale aj zlepšuje kvalitu rozhodovania. Spoločnosti, ktoré implementujú vizuálne analyticke nástroje a ETL -prevodníky, získavajú možnosť rýchlo reagovať na zmeny

Automatizácia podnikových procesov pomocou nástrojov ako n8n, Airflow a NiFi je len prvým krokom k digitálnej vyspelosti. Ďalším krokom je kvalitné ukladanie a správa samotných údajov, ktoré sú základom automatizácie. V ôsmej časti sa podrobne pozrieme na to, ako môžu stavebné spoločnosti vybudovať udržateľnú architektúru ukladania údajov, ktorá prejde od chaosu dokumentov a súborov rôznych formátov k centralizovaným platformám na ukladanie a analýzu.





## VIII ČASŤ

# **UKLADANIE A SPRÁVA ÚDAJOV V STAVEBNÍCTVE**

Časť 8 sa zaoberá modernými technológiami ukladania a správy údajov pre stavebníctvo. Analyzuje efektívne formáty na spracovanie veľkých objemov informácií - od jednoduchých CSV a XLSX až po výkonnejšie Apache Parquet a ORC s podrobným porovnaním ich možností a obmedzení. Rozoberajú sa koncepcie dátových skladov (DWH), dátových jazier) a ich hybridných riešení (Data Lakehouse), ako aj zásady Data Governance) a Data Minimalism). Podrobne sa rozoberajú problémy dátovej bažiny) a stratégie na predchádzanie chaosu v informačných systémoch. Predstavujú sa nové prístupy k práci s údajmi vrátane vektorových databáz a ich uplatnenie pri konštrukcii prostredníctvom koncepcie Bounding Box. Táto časť sa dotýka aj metodík DataOps a VectorOps ako nových štandardov na organizáciu práce s údajmi.

## KAPITOLA 8.1.

# DÁTOVÁ INFRAŠTRUKTÚRA: OD FORMÁTOV UKLADANIA PO DIGITÁLNE ÚLOŽISKÁ

### Dátové atómy: základ efektívneho riadenia informácií

Všetko vo vesmíre sa skladá z najmenších stavebných prvkov - atómov a molekúl a časom sa všetko živé aj neživé nevyhnutne vráti do tohto počiatočného stavu. V prírode tento proces prebieha s ohromujúcou rýchlosťou, ktorú sa snažíme preniesť na procesy riadené človekom.

V lese sa všetky živé organizmy nakoniec premenia na výživnú látku, ktorá slúži ako základ pre nové rastliny. Tieto rastliny sa zasa stávajú potravou pre nové živé organizmy zložené z tých istých atómov, ktoré pred miliónmi rok vytvorili vesmír.

Vo svete podnikania je tiež dôležité rozložiť zložité, viacvrstvové štruktúry na ich najzákladnejšie, minimálne spracované jednotky - podobne ako atómy a molekuly v prírode. To umožňuje atómy údajov efektívne ukladať a spravovať, čím sa z nich stáva bohatý, úrodný základ, ktorý sa stáva kľúčovým zdrojom pre rast analytiky a kvality rozhodovania.



Obr. 8.1-1 Analýza a rozhodovanie sú založené na opakovane používaných údajoch, ktoré boli raz spracované a uložené.

Hudobné skladby sa skladajú z tónov, ktoré sa spájajú do zložitých hudobných diel, zatiaľ čo slová sa vytvárajú zo základnej jednotky, písmena a zvuku. Či už ide o prírodu, vedu, ekonomiku, umenie alebo

technológiu, svet vykazuje pozoruhodnú jednotu a harmóniu v snahe o deštrukciu, štruktúru, cyklický pohyb a tvorbu. Podobne aj procesy v systémoch kalkulácie nákladov sú rozdelené na malé štruktúrované jednotky - položky zdrojov - na úrovni kalkulácie nákladov a harmonogramov. Tieto jednotky sa potom podobne ako poznámky používajú na tvorbu zložitejších kalkulácií a rozvrhov. Rovnaký princíp využívajú systémy počítačom podporovaného projektovania, v ktorých sa komplexné architektonické a inžinierske projekty vytvárajú zo základných prvkov - jednotlivých prvkov a knižničných komponentov, z ktorých sa vytvára kompletný 3D -model projektu komplexnej budovy alebo stavby.

Koncepcia cyklickosti a štruktúry, ktorá je vlastná prírode a vede, sa odráža aj v modernom svete údajov. Tak ako sa v prírode všetko živé vracia k atómom a molekulám, tak aj vo svete moderných nástrojov na spracovanie údajov majú informácie tendenciu vracať sa k svojej najprimitívnejšej forme.

Najmenšie prvky s ich konečnou nedeliteľnosťou sú základnými stavebnými prvkami podnikových procesov. Je dôležité od začiatku starostlivo zvážiť, ako tieto drobné stavebné prvky z rôznych zdrojov zhromažďovať, štruktúrovať (rozložiť na atómy) a ukladať. Organizácia a ukladanie údajov nie je len otázkou ich rozkladu na jednotlivé zložky. Rovnako dôležité je zabezpečiť, aby boli integrované a uložené štruktúrovaným spôsobom, aby sa údaje dali ľahko vyhľadať, analyzovať a použiť na rozhodovanie, kedykoľvek sú potrebné.

Na efektívne spracovanie informácií je potrebné starostlivo zvoliť formát a metódy ukladania údajov - rovnako ako je potrebné pripraviť pôdu na rast stromov. Úložiská údajov musia byť organizované tak, aby sa zabezpečila vysoká kvalita a relevantnosť informácií a aby sa vylúčili nadbytočné alebo irelevantné údaje. Čím lepšie je táto "informačná pôda" štruktúrovaná, tým rýchlejšie a presnejšie môžu používatelia nájsť správne údaje a riešiť analytické problémy.

## **Ukladanie informácií: súbory alebo údaje**

Dátové sklady umožňujú spoločnostiam zhromažďovať a kombinovať informácie z rôznych systémov a vytvárať tak jedno centrum pre ďalšiu analýzu. Zhromaždené historické údaje umožňujú nielen hlbšiu analýzu procesov, ale aj identifikáciu vzorcov, ktoré môžu ovplyvniť výkonnosť podniku.

Povedzme, že spoločnosť pracuje na viacerých projektoch súčasne. Inžinier chce zistiť, koľko betónu sa už vylialo a koľko ho ešte treba nakúpiť. Pri tradičnom prístupe by musel ručne prehľadávať server a otvoriť niekoľko tabuľiek odhadov, porovnať ich s potvrdeniami o vykonaných prácach a skontrolovať aktuálne stavy zásob. To trvá hodiny alebo dokonca dni. Dokonca aj s procesmi ETL a automatickými skriptami zostáva táto úloha čiastočne manuálna: inžinier musí stále ručne určiť cestu k priečinkom alebo konkrétnym súborom na serveri. To znižuje celkový efekt automatizácie, pretože naďalej zaberá cenný pracovný čas.

Pri prechode na správu údajov získava inžinier namiesto práce so súborovým systémom servera prístup k jednotnej štruktúre úložiska, kde sa informácie aktualizujú v reálnom čase. Jediný dotaz - vo forme kódu, SQL -dotazu alebo dokonca volania LLM -agenta - môže okamžite poskytnúť presné údaje o aktuálnych zostatkoch, objemoch vykonanej práce a nadchádzajúcich dodávkach, ak boli údaje

vopred pripravené a skombinované v dátovom skade, kde nie je potrebné blúdiť po priečinkoch, otvárať desiatky súborov a ručne porovnávať hodnoty.

Stavebné spoločnosti dlho používali PDF -dokumenty, DWG -výkresy, RVT -modely a stovky a tisíce tabuľiek v Exceli -tabuľky a iné rôznorodé formáty, ktoré sú uložené v špecifických priečinkoch na serveroch spoločnosti, čo stáže vyhľadávanie informácií, ich kontrolu a analýzu. V dôsledku toho sa súbory, ktoré zostanú po ukončení projektov, najčastejšie presúvajú späť na server do archívnych priečinkov, ktoré sa v budúcnosti prakticky nepoužívajú. Takéto tradičné ukladanie údajov založené na súboroch stráca význam s rastúcim tokom údajov, a to z dôvodu jeho zraniteľnosti voči ľudským chybám.

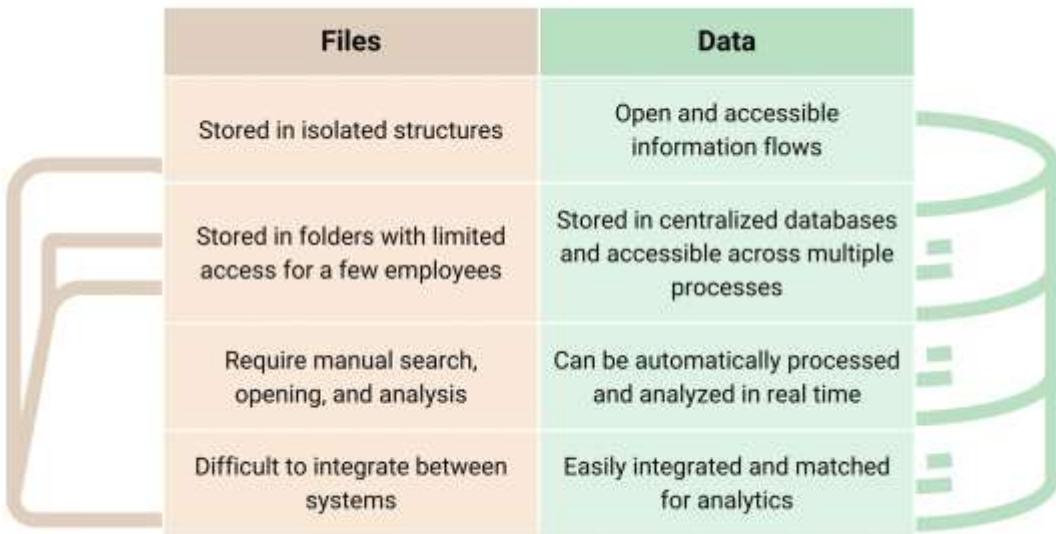
**Súbor** je len izolovaný kontajner, v ktorom sú uložené údaje. Súbory sa vytvárajú pre ľudí, nie pre systémy, takže si vyžadujú manuálne otváranie, čítanie a interpretáciu. Príkladom je tabuľka v programe Excel -tabuľka, dokument PDF -dokument alebo výkres CAD -výkres, ktorý je potrebné osobitne otvoriť v konkrétnom nástroji, aby ste získali prístup k požadovaným informáciám. Bez štruktúrovaného vyhľadávania a spracovania zostanú informácie v ňom nevyužité.

**Údaje** sú zasa strojovo čitateľné informácie, ktoré sa automaticky spájajú, aktualizujú a analyzujú. V jednom dátovom skade (napr. databáze, DWH alebo Data Lake) sú informácie reprezentované vo forme tabuľiek, záznamov a vzťahov. To umožňuje jednotné ukladanie, automatizované dotazy, hodnotové analýzy a reportovanie v reálnom čase.

Používanie údajov namiesto súborov (obr. 8.1-1) eliminuje proces manuálneho vyhľadávania a zjednocuje spracovanie. Spoločnosti, ktoré už takýto prístup zaviedli, získavajú konkurenčnú výhodu vďaka rýchlosťi prístupu k informáciám a schopnosti rýchlo ich integrovať do podnikových procesov.

Prechod od používania súborov k údajom je nevyhnutnou zmenou, ktorá ovplyvní budúcnosť stavebníctva.

Každá spoločnosť v stavebníctve bude stať pred klúčovou voľbou: pokračovať v ukladaní informácií v rôznorodých súboroch a silách, ktoré musia čítať ľudia pomocou špeciálnych programov, alebo ich v prvých fázach spracovania transformovať na štruktúrované údaje a vytvoriť jednotný integrovaný digitálny základ pre automatizované riadenie projektov.



Obrázok 8.1-1 Vývoj toku informácií: od izolovaných súborov k integrovaným údajom.

S prudkým nárastom informácií sú tradičné metódy ukladania a spracovania súborov čoraz menej efektívne. V stavebnictve, rovnako ako v iných odvetviach, už nestačí spoliehať sa na rôznorodé zložky s rôznymi formátm súborov alebo na neprepojené databázy.

Spoločnosti, ktoré si chcú udržať konkurencieschopnosť v digitálnom veku, budú nevyhnutne prechádzať na integrované digitálne platformy, využívať technológie veľkých dát a automatizované analytické systémy.

Prechod od ukladania súborov k pracovným postupom založeným na údajoch si bude vyžadovať prehodnotenie prístupov k správe informácií a vedomý výber formátov vhodných na ďalšiu integráciu do centralizovaných úložísk. Tento výber bude určovať, ako efektívne možno údaje spracovať, ako rýchlo k nim možno pristupovať a ako ľahko ich možno integrovať do digitálnych procesov spoločnosti.

## Ukladanie veľkých objemov údajov: Analýza populárnych formátov a ich efektívnosti

Formáty úložísk zohrávajú klúčovú úlohu pri škálovateľnosti, spoľahlivosti a výkonnosti analytickej infraštruktúry. Na analýzu a spracovanie údajov - napríklad filtrovanie, zoskupovanie a agregáciu - sme v našich príkladoch použili Pandas DataFrame - populárnu štruktúru na prácu s údajmi v pamäti RAM.

Pandas DataFrame však nemá vlastný formát ukladania, takže po dokončení spracovania sa údaje exportujú do jedného z externých formátov - najčastejšie CSV alebo XLSX. Tieto tabuľkové formáty sa ľahko vymieňajú a sú kompatibilné s väčšinou externých systémov, ale majú niekoľko obmedzení: nízku efektívnosť ukladania, nedostatočnú kompresiu a slabú podporu verzíí:

- **CSV** (Comma-Separated Values - hodnoty oddelené čiarkou): jednoduchý textový formát, ktorý je široko podporovaný rôznymi platformami a nástrojmi. Ľahko sa používa, ale nepodporuje

zložité typy údajov a kompresiu.

- **XLSX** (Excel Open XML Spreadsheet): formát súboru Microsoft Excel, ktorý podporuje sofistikované funkcie, ako sú vzorce, grafy a štýly. Hoci je užitočný na manuálnu analýzu a vizualizáciu údajov, nie je optimalizovaný na rozsiahle spracovanie údajov.

Okrem populárnych tabuľkových formátov XLSX a CSV, existuje niekoľko populárnych formátov na efektívne ukladanie štruktúrovaných údajov (obrázok 8.1-2), pričom každý z nich má jedinečné výhody v závislosti od špecifických požiadaviek na ukladanie a analýzu údajov:

- **Apache Parquet**: formát stĺpcového ukladania údajov optimalizovaný na použitie v systémoch na analýzu údajov. Ponúka účinné schémy kompresie a kódovania údajov, vďaka čomu je ideálny pre komplexné dátové štruktúry a spracovanie veľkých objemov údajov.
- **Apache ORC** (Optimised Row Columnar): Podobne ako Parquet, ORC poskytuje vysokú kompresiu a efektívne ukladanie údajov. Je optimalizovaný na náročné operácie čítania a je vhodný na ukladanie dátových jazier.
- **JSON** (JavaScript Object Notation): JSON síce nie je tak efektívny z hľadiska ukladania údajov v porovnaní s binárnymi formátm, ako sú Parquet alebo ORC, ale je veľmi prístupný a ľahko sa s ním pracuje, takže je ideálny pre skripty, pri ktorých je dôležitá čitateľnosť a kompatibilita s webom.
- **Feather**: rýchly, ľahký a ľahko použiteľný binárny stĺpcový formát ukladania údajov orientovaný na analýzu. Je navrhnutý na efektívny prenos údajov medzi jazykmi Python (Pandas) a R, vďaka čomu je vynikajúcou voľbou pre projekty zahŕňajúce tieto programovacie prostredia.
- **HDF5** (Hierarchický dátový formát verzia 5): určený na ukladanie a organizáciu veľkého množstva údajov. Podporuje širokú škálu dátových typov a je vhodný na prácu so zložitými súbormi údajov. HDF5 je obzvlášť oblúbený vo vedeckých výpočtoch vďaka svojej schopnosti efektívne ukladať a sprístupňovať veľké súbory údajov.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame	
	Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
	Compression	Built-in	None	High	Built-in	None (in-memory)
	Performance	Low	Medium	High	High	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Medium	Medium	Low
	Data Type Support	Limited	Very limited	Extended	Extended	Extended
	Scalability	Low	Low	High	High	Medium (memory limited)

Obr. 8.1-2 Porovnanie formátov údajov, ktoré ukazujú hlavné rozdiely v aspektoch ukladania a spracovania.

Na vykonanie porovnávacej analýzy formátov použitých vo fáze Load procesu ETL bola vytvorená tabuľka zobrazujúca veľkosti súborov a časy čítania (obr. 8.1-3). V štúdii boli použité súbory s identickými údajmi: tabuľka obsahovala 10 000 riadkov a 10 stĺpcov vyplnených náhodnými hodnotami.

V štúdii sú zahrnuté tieto formáty ukladania: CSV, Parquet, XLSX a HDF5, ako aj ich komprimované verzie v archívoch ZIP. Surové údaje boli vytvorené pomocou knižnice NumPy a reprezentované ako štruktúra Pandas DataFrame. Proces testovania pozostával z nasledujúcich krokov:

- Ukladanie súborov: Dátový rámec sa ukladá v štyroch rôznych formátoch: CSV, Parquet, XLSX a HDF5. Každý formát má jedinečné vlastnosti v spôsobe ukladania údajov, ktoré ovplyvňujú veľkosť súboru a rýchlosť čítania.
- Komprezia súborov ZIP: na analýzu účinnosti štandardnej kompresie bol každý súbor ďalej komprimovaný do archívu ZIP.
- Čítanie súborov (ETL - Load): čas čítania sa meral pre každý súbor po jeho rozbalení zo ZIP. To umožňuje odhadnúť rýchlosť prístupu k údajom po rozbalení z archívu.

Je dôležité poznamenať, že Pandas DataFrame neboli použití priamo pri analýze veľkosti alebo času čítania, pretože nepredstavuje samostatný formát ukladania. Slúžil len ako medzistruktúra na generovanie a následné ukladanie údajov do rôznych formátov.



Obr. 8.1-3 Porovnanie pamäťových formátov podľa veľkosti a rýchlosť čítania.

Súbory CSV a HDF5 vykazujú (obrázok 8.1-3) vysokú účinnosť kompresie, ktorá výrazne znižuje ich veľkosť pri zabalení do ZIP, čo môže byť užitočné najmä v scenároch vyžadujúcich optimalizáciu ukladania. Na druhej strane súbory XLSX sú prakticky nekomprimovateľné a ich veľkosť v ZIP zostáva porovnatelná s originálom, čo ich robí menej výhodnými na použitie pri veľkých objemoch údajov alebo v prostrediach, kde je dôležitá rýchlosť prístupu k údajom. Okrem toho je čas čítania súborov XLSX v porovnaní s inými formátmi výrazne vyšší, čo ich robí menej výhodnými pre rýchle operácie čítania údajov. Apache Parquet preukázal vysoký výkon pre analytické úlohy a veľké objemy údajov vďaka svojej stĺpcovej štruktúre.

### Optimalizácia ukladania pomocou Apache Parquet

Jedným z populárnych formátov na ukladanie a spracovanie veľkých dát je Apache Parquet. Tento formát je špeciálne navrhnutý na stĺpcové ukladanie (podobne ako Pandas), čo umožňuje výrazne znižiť pamäťovú náročnosť a zvýšiť rýchlosť analytických dotazov. Na rozdiel od tradičných formátov, ako sú CSV a XLSX, Parquet podporuje natívnu kompresiu a je optimalizovaný pre systémy veľkých dát vrátane Sparku, Hadoopu a clouдовých úložísk.

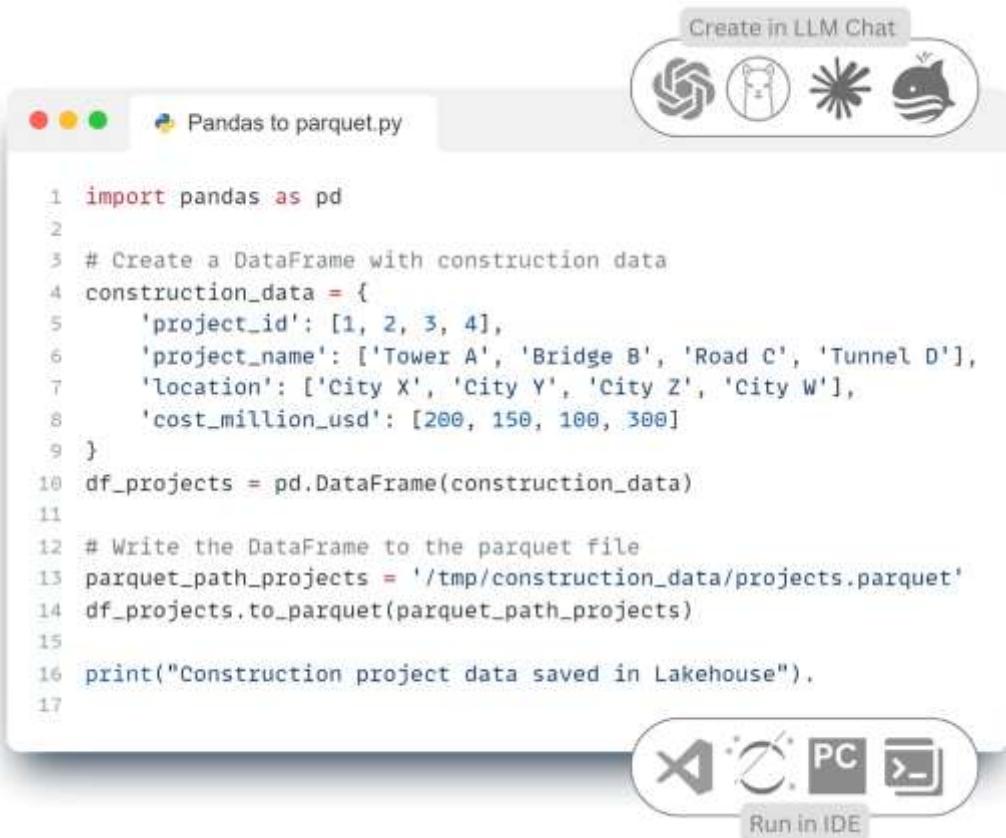
Medzi kľúčové vlastnosti systému Parquet patrí podpora kompresie a kódovania údajov, ktorá výrazne znižuje veľkosť úložiska a urýchľuje operácie čítania údajov tým, že pracuje priamo s požadovanými stĺpcami, a nie so všetkými riadkami údajov.

Na ilustračný príklad toho, ako ľahko možno získať potrebný kód na konverziu údajov do Apache Parquet, použime LLM.

- Pošlite textovú požiadavku do chatu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Napíšte kód na uloženie údajov z Pandas DataFrame do Apache Parquet. ↵

- Odpoveď LLM :



The screenshot shows a user interface for generating code. At the top, there's a button labeled "Create in LLM Chat" and a row of icons representing different AI models: ChatGPT, LLaMA, Mistral, DeepSeek, Grok, Claude, and Qwen. Below this is a code editor window with the title "Pandas to parquet.py". The code itself is:

```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

At the bottom of the code editor, there's a "Run in IDE" button with icons for VS Code, PyCharm, and others.

**Obrázok 8.1-4 Prenos údajov Dataframe z pamäte RAM do formátu Apache Parquet efektívneho z hľadiska ukladania pomocou niekoľkých riadkov jazyka Python .**

Ďalší príklad: simulujme proces ETL s údajmi uloženými vo formáte Parquet na filtrovanie projektov podľa určitej hodnoty jedného z atribútov "cost\_million\_usd" (obr. 8.1-4).

- Počas pokračovania chatu pošlite textovú žiadosť na adresu LLM:

Napíšte kód, v ktorom chceme filtrovať údaje v tabuľke a uložiť len tie projekty (riadky tabuľky) z údajov Apache Parquet, ktorých náklady (parameter cost\_million\_usd) presahujú 150 miliónov USD. ↴

💡 Odpoveď LLM :



```
# Loading construction project data
df_loaded_projects = pd.read_parquet(parquet_path_projects)

# Filtering out projects with a value greater than $150 million
df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]

# Saving the filtered data to a new Parquet file
filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
df_filtered_projects.to_parquet(filtered_parquet_path_projects)

print("Filtered construction projects data saved to Lakehouse.")
```

The screenshot shows a code editor window titled "ETL parquet.py". The code itself is a Python script for performing ETL operations on Apache Parquet files. It starts by loading data from a Parquet file, then filters out projects where the cost is greater than 150 million USD, and finally saves the filtered data back to a new Parquet file. Below the code editor, there are several icons: a neural network icon, a cat icon, a starburst icon, a whale icon, a "Run in IDE" button, and a "PC" icon.

Obrázok 8.1-5 Proces ETL pri práci s údajmi vo formáte Apache Parquet vyzerá rovnako ako pri iných štruktúrovaných formátoch.

Používanie formátu Parquet (v porovnaní s formátmi XLSX, CSV atď.) výrazne znižuje množstvo uložených informácií a urýchľuje vyhľadávanie. Vďaka tomu je vynikajúci na ukladanie aj analýzu údajov. Formát Parquet sa integruje s rôznymi systémami spracovania a poskytuje efektívny prístup v hybridných architektúrach.

Efektívny formát ukladania je však len jedným prvkom kompletného zájitku z dát. Na vytvorenie udržateľného a škálovateľného prostredia je potrebná dobre navrhnutá architektúra správy údajov. Túto funkciu plnia systémy triedy DWH (Data Warehouse). Poskytujú agregáciu údajov z heterogénnych zdrojov, transparentnosť obchodných procesov a možnosť komplexnej analýzy pomocou nástrojov BI a algoritmov strojového učenia.

## DWH: Dátový sklad dátové sklady

Tak ako je formát Parquet optimalizovaný na efektívne ukladanie veľkého množstva informácií, tak je dátový sklad optimalizovaný na integráciu a štruktúrovanie údajov na podporu analytiky, prognózovania a rozhodovania manažmentu.

V dnešných spoločnostiach pochádzajú údaje z mnohých rôznorodých zdrojov: systémov ERP, CAFM, CPM, CRM, účtovníctva a skladového hospodárstva, digitálnych modelov CAD budov, senzorov IoT a ďalších riešení. Na získanie uceleného obrazu nestačí údaje len zhromažďovať - je potrebné ich

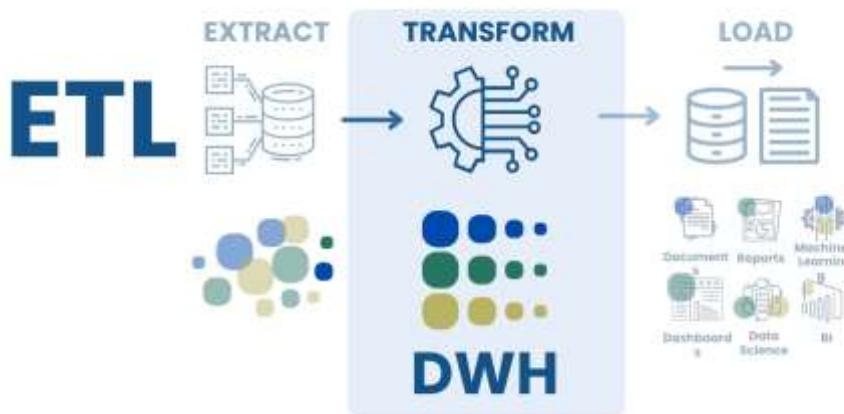
organizovať, štandardizovať a centralizovať v jednom úložisku. Práve to robí DWH - centralizovaný systém úložiska, ktorý umožňuje zhromažďovať informácie z rôznych zdrojov, štruktúrovať ich a sprístupniť na analýzu a strategické riadenie.

**DWH (Data Warehouse)** je centralizovaný systém dátového skladu, ktorý zhromažďuje informácie z viacerých zdrojov, štruktúruje ich a sprístupňuje na analytické účely a reportovanie.

V mnohých spoločnostiach sú údaje rozptýlené v rôznych systémoch, o ktorých sme hovorili v prvých častiach knihy (obrázok 1.2-4). DWH integruje tieto zdroje, čím zabezpečuje úplnú prehľadnosť a spoľahlivosť informácií. Dátový sklad DWH je špecializovaná databáza (veľká databáza), ktorá zhromažďuje, spracováva a uchováva údaje z viacerých zdrojov. Hlavné charakteristiky DWH sú:

- **Použitie ETL -procesov** (Extract, Transform, Load) - extrakcia údajov zo zdrojov, ich čistenie, transformácia, načítanie do úložiska a automatizácia týchto procesov, o ktorých sa hovorilo v siedmej časti knihy.
- **Granularita údajov** - údaje v DWH môžu byť uložené v agregovanej forme (súhrnné správy) aj v granulovanej forme (nespracované údaje). Od roku 2024 začali o granulovaných údajoch hovoriť práve dodávateelia CAD- [125], čo možno naznačuje, že odvetvie sa pripravuje na prechod na špecializované cloudové úložisko na spracovanie údajov digitálneho modelu budovy.
- **Podpora analytických a prediktívnych** - dátových skladov poskytuje základ pre nástroje BI, Big Data -analýzu a strojové učenie.

DWH slúži ako základ pre business intelligence, umožňuje analýzu klúčových ukazovateľov výkonnosti, prognózovanie predaja, nákupu a nákladov a automatizované reportovanie a vizualizáciu údajov (obrázok 8.1-6).



Obr. 8.1-6 V procese ETL môže DWH fungovať ako centrálné úložisko, kde údaje získané z rôznych systémov prechádzajú fázami transformácie a vykladania.

DWH zohráva klúčovú úlohu pri integrácii, čistení a štruktúrovaní informácií, čím poskytuje pevný

základ pre business intelligence a rozhodovacie procesy. V dnešnom prostredí, kde objemy údajov rýchlo rastú a zdroje údajov sú čoraz rozmanitejšie, si však tradičný prístup DWH k ukladaniu informácií často vyžaduje rozšírenie v podobe ELT a Data Lake

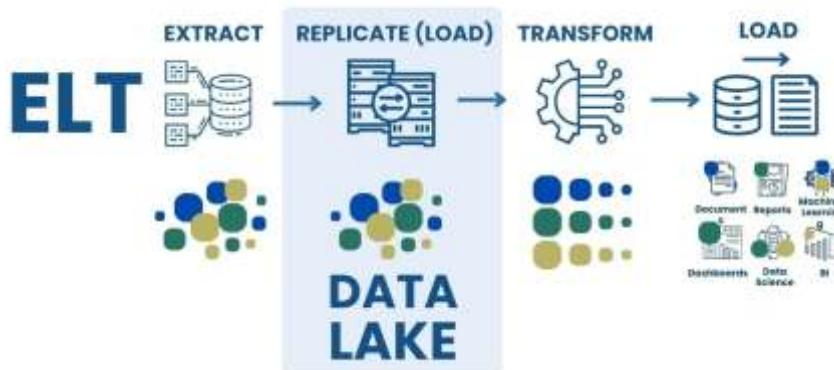
## Data Lake - vývoj ETL na ELT: od tradičného čistenia k flexibilnému spracovaniu

Klasické DWH - dátové sklady, navrhnuté na ukladanie štruktúrovaných údajov vo formáte optimalizovanom na analytické dotazy, narazili na obmedzenia pri spracovaní neštruktúrovaných údajov a škálovateľnosti. Ako odpoveď na tieto výzvy sa objavili dátové jazerá), ktoré ponúkajú flexibilné ukladanie veľkého množstva heterogénnych údajov.

**Data Lake** ponúka alternatívny prístup DWH, ktorý umožňuje pracovať s neštruktúrovanými, pološtruktúrovanými a nespracovanými údajmi bez predchádzajúcej pevnej schémy. Tento spôsob ukladania je často relevantný pre spracovanie údajov v reálnom čase, strojové učenie a pokročilú analýzu. Na rozdiel od DWH, ktorý údaje pred načítaním štruktúruje a agreguje, Data Lake umožňuje ukladať informácie v ich surovej podobe, čím poskytuje flexibilitu a škálovateľnosť

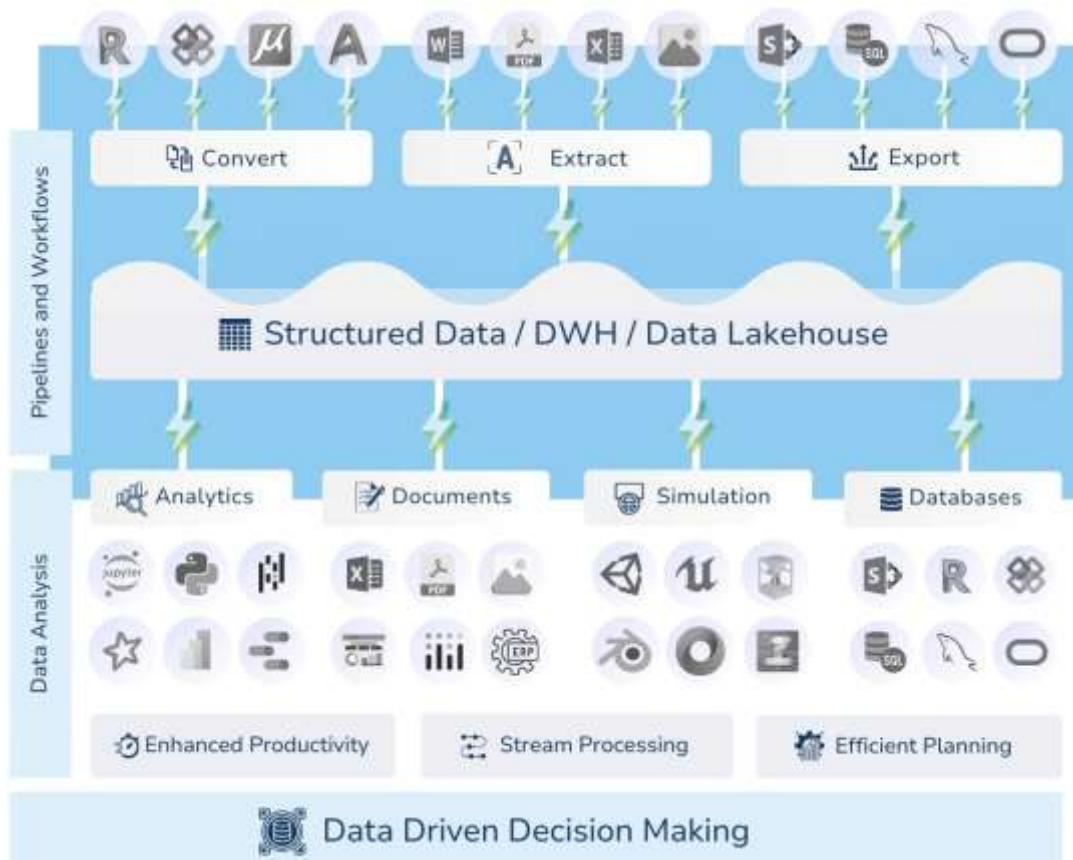
Práve frustrácia z tradičných dátových skladov (RDBMS, DWH) a záujem o "veľké dáta" viedli k vzniku dátových jazier, kde sa namiesto zložitého ETL údaje jednoducho nahrávajú do voľne štruktúrovaného úložiska a spracúvajú sa vo fáze analýzy:

- V tradičných dátových skladoch sa údaje pred vložením do skladu zvyčajne predbežne spracúvajú, transformujú a čistia (ETL - Extract, Transform, Load) (obrázok 8.1-6). To znamená, že údaje sú štruktúrované a optimalizované pre konkrétné budúce analytické a reportovacie úlohy. Dôraz sa kladie na zachovanie vysokej výkonnosti dopytovania a integrity údajov. Tento prístup však môže byť nákladný a menej flexibilný z hľadiska integrácie nových typov údajov a rýchlo sa meniacich dátových schém.
- Na druhej strane sú dátové jazerá určené na ukladanie veľkého množstva nespracovaných údajov v pôvodnom formáte (obrázok 8.1-7). Proces ETL (Extract, Transform, Load), sa nahrádza procesom ELT (Extract, Load, Transform), pri ktorom sa údaje najprv načítajú do dátového skladu "tak, ako sú", až potom sa môžu podľa potreby transformovať a analyzovať. To poskytuje väčšiu flexibilitu a možnosť ukladať heterogénne údaje vrátane neštruktúrovaných údajov, ako sú text, obrázky a protokoly.



Obr. 8.1-7 Na rozdiel od ETL, Data Lake využíva ELT, v ktorom sa informácie najprv nahrajú v "surovej" forme a transformácia sa vykonáva vo fáze nahrávania.

Tradičné dátové sklady sa zameriavajú na predbežné spracovanie údajov s cieľom zabezpečiť vysoký výkon dotazov, zatiaľ čo dátové jazerá uprednostňujú flexibilitu: ukladajú surové údaje a transformujú ich podľa potreby (obrázok 8.1-8).



Obr. 8.1-8 Moderné koncepte ukladania dát sú zamerané na ukladanie a spracovanie všetkých typov dát na účely rozhodovania.

Napriek všetkým výhodám však dátové jazerá nie sú bez nevýhod. Absencia prísnej štruktúry a zložitosť správy informácií môže viesť k chaosu, v ktorom sa údaje duplikujú, navzájom si odporuju-

alebo sa stávajú irelevantnými. Okrem toho si vyhľadávanie a analýza údajov v takomto úložisku vyžaduje značné úsilie, najmä ak ide o heterogénne informácie. Na prekonanie týchto obmedzení a spojenie najlepších vlastností tradičných dátových skladov a dátových jazier bola vyvinutá architektúra Data Lakehouse.

## Architektúra dátového skladu: synergia skladov a dátových jazier

Na spojenie najlepších vlastností DWH (štruktúrovaná, spravovateľná, vysoko výkonná analytika) a Data Lake (škálovateľnosť, spracovanie heterogénnych údajov) bol vyvinutý prístup Data Lakehouse. Táto architektúra spája flexibilitu dátových jazier s výkonnými nástrojmi na spracovanie a správu údajov typickými pre tradičné skladby, pričom dosahuje rovnováhu medzi ukladaním, analýzou a strojovým učením. Data Lakehouse je syntézou dátových jazier a dátových skladov, pričom spája flexibilitu a škálovateľnosť prvých s manažovateľnosťou a optimalizáciou dotazov druhých.

**Dátové jazero** je architektonický prístup, ktorý sa snaží spojiť flexibilitu a škálovateľnosť dátových jazier so spravovateľnosťou a výkonnosťou dotazov dátových skladov (obrázok 8.1-9).

Medzi kľúčové vlastnosti Data Lakehouse patria:

- **Otvorený formát ukladania údajov:** používanie otvorených formátov na ukladanie údajov, ako je napríklad Apache Parquet, zabezpečuje efektívnosť a optimalizáciu dotazov.
- **Schéma len na čítanie:** na rozdiel od tradičného prístupu schémy len na zápis v DWH, Lakehouse podporuje schému len na čítanie, ktorá umožňuje väčšiu flexibilitu pri správe štruktúry údajov.
- **Flexibilný a škálovateľný:** podporuje ukladanie a analýzu štruktúrovaných a neštruktúrovaných údajov a poskytuje vysoký výkon dotazov vďaka optimalizácii na úrovni úložiska.

Data Lakehouse ponúka kompromisné riešenie, ktoré spája výhody oboch prístupov, takže je ideálne pre moderné analytické úlohy, ktoré vyžadujú flexibilitu pri spracovaní údajov.



Obr. 8.1-9 Data Lakehouse je nová generácia systémov na ukladanie dát navrhnutá tak, aby spĺňala komplexné a neustále sa meniace požiadavky.

Myšlienka moderných dátových skladov sa zdá byť jednoduchá: ak sú všetky údaje na jednom mieste, ľahšie sa analyzujú. V praxi však nie je všetko také hladké. Predstavte si, že sa spoločnosť rozhodne úplne opustiť bežné účtovné a riadiace systémy (ERP, PMIS, CAFM alebo iné) a nahradí ich jedným obrovským dátovým jazerom, ku ktorému majú prístup všetci. Čo sa stane? S najväčšou pravdepodobnosťou nastane chaos: údaje budú duplicitné, protichodné a kritické informácie sa stratia alebo poškodia. Aj keď sa dátové jazero používa len na analytické účely, bez správnej správy bude väčne ohrozené:

- Dáta sa ľahko chápu: v bežných systémoch majú dátá jasnú štruktúru, ale v jazere sú len obrovským nahromadením súborov a tabuliek. Ak chce niečo nájsť, musí zistiť, za čo sú zodpovedné jednotlivé riadky a stĺpce.
- Údaje môžu byť nepresné: ak je na jednom mieste uložených mnoho verzií tých istých informácií, je ľahké zistiť, ktorá verzia je aktuálna. V dôsledku toho sa rozhodnutia prijímajú na základe neaktuálnych alebo chybných údajov.
- Je náročné pripraviť údaje na prácu: údaje musia byť nielen uložené, ale aj prezentované vo vhodnej forme - vo forme správ, grafov, tabuliek. V tradičných systémoch sa to robí automaticky, ale v dátových jazerach si to vyžaduje ďalšie spracovanie.

V dôsledku toho má každá koncepcia dátového skladu svoje vlastné charakteristiky, prístupy k spracovaniu a obchodné aplikácie. Tradičné databázy sa zameriavajú na transakčné operácie, dátové sklad (DWH) poskytujú štruktúru pre analytiku, dátové jazerá (Data Lake) uchovávajú informácie v súrovej forme a hybridné sklad (Data Lakehouse) kombinujú výhody DWH a Data Lake (obrázok 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

Obr. 8.1-10 DWH, Data Lake a Data Lakehouse: hlavné rozdiely v typoch údajov, scenároch použitia, metódach spracovania a prístupoch k ukladaniu.

Výber architektúry úložiska je zložitý proces, ktorý závisí od obchodných potrieb, objemu informácií a požiadaviek na analýzu. Každé riešenie má svoje výhody a nevýhody: DWH poskytuje štruktúru, Data Lake poskytuje flexibilitu a Lakehouse poskytuje rovnováhu medzi nimi. Organizácie sú len zriedkakedy obmedzené na jednu dátovú architektúru.

Bez ohľadu na zvolenú architektúru sú automatizované systémy správy údajov výrazne lepšie ako manuálne metódy. Minimalizujú ľudské chyby, urýchľujú spracovanie informácií a zabezpečujú transparentnosť a sledovateľnosť údajov vo všetkých fázach obchodných procesov.

Zatiaľ čo centralizované dátové sklady sa už stali priemyselným štandardom v mnohých oblastiach hospodárstva, situácia v stavebnictve je stále rozrieštená. Údaje tu sú distribuované na rôznych platformách (CDE, PMIS, ERP atď.), čo sťaže vytvorenie jednotného obrazu o dianí a vyžaduje architektúry schopné spojiť tieto zdroje do uceleného, analyticky využiteľného digitálneho prostredia.

## CDE, PMIS, ERP alebo DWH a dátové jazero

Niekteré stavebné a inžinierske spoločnosti už používajú koncept spoločného dátového prostredia (CDE) podľa normy ISO 19650. CDE v podstate plní rovnaké funkcie ako dátový sklad (DWH) v iných odvetviach: centralizuje informácie, zabezpečuje kontrolu verzií, poskytuje prístup k overeným informáciám.

**Spoločné dátové prostredie (CDE)** je centralizovaný digitálny priestor, ktorý sa používa na správu, ukladanie, zdieľanie a spoluprácu na projektových informáciách počas všetkých fáz životného cyklu zariadenia. CDE sa často implementuje pomocou clouдовých technológií a integruje sa so systémami CAD (BIM).

Finančné, maloobchodné, logistické a priemyselné odvetvia už desaťročia používajú centralizované systémy správy údajov, ktoré kombinujú informácie z rôznych zdrojov, kontrolujú ich relevantnosť a poskytujú analytické služby. CDE posúva tieto princípy ďalej tým, že ich prispôsobuje výzvam v oblasti navrhovania a riadenia životného cyklu budov.

Podobne ako DWH, CDE štruktúruje údaje, zachytáva zmeny a poskytuje jediný prístupový bod k overeným informáciám. S prechodom na cloud a integráciou s analytickými nástrojmi sú rozdiely medzi nimi čoraz menej zjavné. Ak sa k CDE pridajú granulárne údaje, o ktorých koncepcii diskutujú dodávatelia CAD od roku 2023[93, 125], možno vidieť ešte viac paralel s klasickým DWH.

V predchádzajúcej kapitole "Stavebný ERP a PMIS systémy" sme sa už zaoberali PMIS (Project Management Information System) a ERP (Enterprise Resource Planning). V stavebných projektoch CDE a PMIS spolupracujú: CDE slúži ako úložisko údajov vrátane výkresov, modelov a projektovej dokumentácie, zatiaľ čo PMIS riadi procesy, ako je kontrola termínov, úloh, zdrojov a rozpočtov.

ERP, ktorý zodpovedá za riadenie podniku ako celku (financie, obstarávanie, personál, výroba), sa môže integrovať so systémom PMIS, ktorý zabezpečuje kontrolu nákladov a rozpočtu na úrovni podniku. Na účely analýzy a vykazovania možno DWH použiť na zhromažďovanie, štruktúrovanie a agregovanie údajov z CDE, PMIS a ERP na vyhodnotenie finančných KPI (návratnosť investícii) a identifikáciu modelov. Dátové jazero (DL) môže zasa dopĺňať DWH ukladaním nespracovaných a neštruktúrovaných údajov (napr. protokolov, údajov zo senzorov, obrázkov). Tieto údaje možno spracovať a načítať do DWH na ďalšiu analýzu.

CDE a PMIS sa teda zameriavajú na riadenie projektov, ERP na obchodné procesy a DWH a Data Lake na analytiku a dátové operácie.

Pri porovnaní systémov CDE, PMIS a ERP so systémami DWH a Data Lake možno pozorovať významné rozdiely v z hľadiska nezávislosti od dodávateľa, nákladov, flexibility integrácie, nezávislosti od údajov, rýchlosťi prispôsobenia sa zmenám a analytických schopností (obrázok 8.1-11). Tradičné systémy, ako sú CDE, PMIS a ERP, sú často viazané na riešenia a štandardy konkrétnych dodávateľov, čo ich robí menej flexibilnými a zvyšuje ich náklady v dôsledku licencíí a podpory. Okrem toho sú údaje v takýchto systémoch často zapuzdrené v proprietárnych, uzavretých formátoch, čo obmedzuje ich použitie a analýzu.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Obr. 8.1-11 DWH a Data Lake ponúkajú väčšiu flexibilitu a nezávislosť od údajov ako systémy ako CDE, PMIS a ERP.

Naproti tomu DWH a Data Lake poskytujú väčšiu flexibilitu pri integrácii s rôznymi zdrojmi údajov a ich využívanie otvorených technológií a platforiem pomáha znížiť celkové náklady na vlastníctvo. DWH a Data Lake navyše podporujú širokú škálu analytických nástrojov, čo zvyšuje možnosti analýzy a správy.

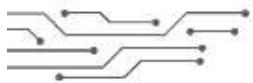
S rozvojom nástrojov reverzného inžinierstva pre formáty CAD a prístupom k databázam aplikácií CAD sa čoraz naliehavejšie kladie otázka: nakoľko je opodstatnené naďalej používať uzavreté, izolované platformy, ak musia byť údaje o návrhu dostupné širokému okruhu odborníkov pracujúcich v desiatkach dodávateľských a projektových organizácií?

Táto technologická závislosť od konkrétnego dodávateľa môže výrazne obmedziť flexibilitu správy údajov, spomaliť reakcie na zmeny v projekte a brániť efektívnej spolupráci medzi účastníkmi.

Tradičné prístupy k správe údajov - vrátane DWH, Data Lake, CDE a PMIS - sa zameriavajú predovšetkým na ukladanie, štruktúrovanie a spracovanie informácií. S rozvojom umelej inteligencie a strojového učenia však rastie potreba nových spôsobov organizácie údajov, ktoré nielen agregujú, ale aj odhalujú komplexné vzťahy, nachádzajú skryté vzory a poskytujú okamžitý prístup k najdôležitejším informáciám.

Osobitnú úlohu v tomto smere začínajú zohrávať vektorové databázy - nový typ úložiska

optimalizovaný pre vysokodimenzionálne osadenia.



## KAPITOLA 8.2.

# SPRÁVA DÁTOVÉHO SKLADU A PREVENCIA CHAOSU

### Vektorové databázy a ohraničujúci rámček

Vektorové databázy sú novou triedou úložísk, ktoré neukladajú len údaje, ale umožňujú vyhľadávanie podľa významu, porovnávanie objektov podľa sémantickej blízkosti a vytváranie inteligentných systémov: od odporúčaní až po automatickú analýzu a generovanie kontextu. Na rozdiel od tradičných databáz, ktoré sa zameriavajú na presné zhody, vektorové databázy vyhľadávajú podobné objekty na základe atribútov - aj keď neexistuje presná zhoda

**Vektorová** databáza je špecializovaný typ databázy, ktorá ukladá údaje ako viacrozmerné vektory, z ktorých každý predstavuje určité charakteristiky alebo vlastnosti. Tieto vektory môžu mať rôzny počet dimenzií v závislosti od zložitosti údajov (v jednom prípade to môže byť niekoľko dimenzií a v inom — tisíc).

Hlavnou výhodou vektorových databáz je vyhľadávanie podľa sémantickej relevantnosti, a nie podľa presnej zhody hodnôt. Namiesto SQL- a Pandas -dotazov s filtrovami "rovná sa" alebo "obsahuje" sa používa vyhľadávanie najbližších susedov (k-NN) (o k-NN budeme hovoriť viac v ďalšej časti knihy) v priestore príznakov.

S rozvojom LLM (Large Language Models) a generatívnych modelov sa interakcia s databázami začína meniť. V súčasnosti je možné vyhľadávať údaje v prirodzenom jazyku, získavať sémantické vyhľadávanie v dokumentoch, automaticky extrahovať klúčové výrazy a vytvárať kontextové vzťahy medzi objektmi - to všetko bez potreby znalosti jazyka SQL alebo znalosti štruktúry tabuľiek. Podrobnejšie sme sa tomu venovali v časti "LLM a ich úloha pri spracovaní údajov a obchodných procesoch".

Je však dôležité uvedomiť si, že LLM neštrukturalizuje a neusporadúva informácie automaticky. Model len prepláva údajmi a nájde najrelevantnejšiu časť údajov na základe kontextu dopytu. Ak údaje neboli vopred očistené alebo transformované, hĺbkové vyhľadávanie bude ako snaha nájsť odpoved' v digitálnom "smeti" - môže fungovať, ale kvalita výsledkov bude nižšia. Ideálne je, ak sa údaje dajú štruktúrovať (napr. preložiť dokumenty do jazyka Markdown) a načítať do vektorovej databázy. Tým sa výrazne zvýši presnosť a relevantnosť výstupu.

Spočiatku sa vektorové databázy používali v strojovom učení, ale dnes nachádzajú čoraz viac aplikácií aj mimo neho - vo vyhľadávačoch, personalizácii obsahu a intelligentnej analytike.

Jedným z najzreteľnejších príkladov vektorového prístupu v konštrukcii je ohraničujúci box (ohraničujúci rovnobežník). Je to geometrická konštrukcia, ktorá opisuje hranice objektu v trojrozmernom priestore. Bounding Box je definovaný minimálnymi a maximálnymi súradnicami X, Y a Z, ktoré tvoria "škatuľu" okolo objektu. Táto metóda umožňuje odhadnúť veľkosť a umiestnenie prvku

bez toho, aby ste museli analyzovať celú geometriu.

Každý Bounding Box možno reprezentovať ako vektor vo viacozmernom priestore: napríklad [x, y, z, šírka, výška, hĺbka] - už 6 rozmerov (obr. 8.2-1)



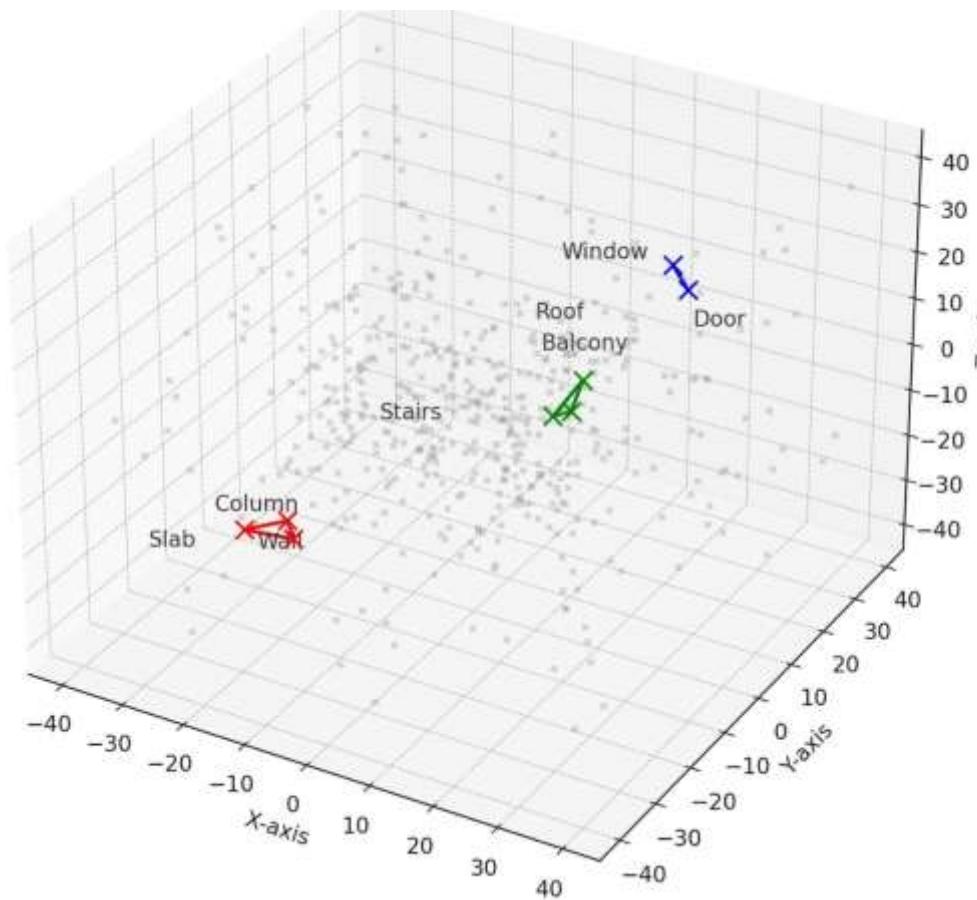
	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

Obr. 8.2-1 Bounding Box -informácie o súradničach prvkov a ich umiestnenie v modeli projektu je analogické vektorovej databáze.

Táto reprezentácia údajov uľahčuje mnohé úlohy vrátane kontroly priesecníkov medzi objektmi, plánovania priestorového rozloženia prvkov budovy a vykonávania automatizovaných výpočtov. Bounding Box môže slúžiť ako most medzi komplexnými 3D modelmi a tradičnými vektorovými databázami, čo vám umožní efektívne využívať výhody oboch prístupov pri architektonickom a inžinierskom modelovaní

Bounding Box je "vektorizácia geometrie" a embedding (spôsob transformácie niečoho abstraktného) je "vektorizácia významu". Oba prístupy umožňujú prejsť od manuálneho vyhľadávania k inteligentnému vyhľadávaniu, či už ide o 3D objekty v modeli projektu alebo pojmy v texte.

Vyhľadávanie objektov v projekte (napríklad "nájdi všetky okná so šírkou > 1,5 m") je podobné vyhľadávaniu najbližších susedov (k-NN) vo vektorovej databáze, kde kritériá definujú "zónu" v priestore prvkov. (o vyhľadávaní k-NN najbližších susedov budeme viac hovoriť v ďalšej časti o strojovom učení) (obr. 8.2-2). Ak k atribútom ohraničenia pridáme ďalšie parametre (materiál, hmotnosť, čas výroby), tabuľka sa zmení na vysokorozmerný vektor, kde každý atribút predstavuje novú dimenziu. To je bližšie k moderným vektorovým bázam, kde sa dimenzie počítajú na stovky alebo tisíce (napr. vkladanie z neurónových sietí).



Obr. 8.2-2 Vyhľadávanie objektov v projekte pomocou vektorových databáz.

Prístup použitý v Bounding Box, je použiteľný nielen na geometrické objekty, ale aj na analýzu textu a jazyka. Vektorové reprezentácie údajov sa už aktívne používajú pri spracovaní prirozeného jazyka (NLP). Tak ako objekty v stavebnom projekte možno zoskupiť podľa ich priestorovej blízkosti (obrázok 8.2-2), slová v texte možno analyzovať podľa ich sémantickej a kontextovej blízkosti.

Napríklad slová "architekt", "konštrukcia", "dizajn" budú vo vektorovom priestore vedľa seba, pretože majú podobný význam. V LLM tento mechanizmus umožňuje automatickú, nevyžadujúcu manuálnu kategorizáciu:

- Identifikovať tému textu
- Vykonávanie sémantickej vyhľadávania v obsahu dokumentov
- Generovanie automatických anotácií a zhrnutí textu
- Nájsť synonymá a súvisiace pojmy

Vektorové databázy umožňujú analyzovať text a nájsť v ňom súvisiace výrazy rovnakým spôsobom, ako vám Bounding Box pomáha analyzovať priestorové objekty v 3D -modeloch. Príklad Bounding Box prvkov projektu pomáha pochopiť, že vektorová reprezentácia nie je čisto "umelý" koncept z ML, ale prirozený spôsob štruktúrovania údajov na riešenie aplikáčnych problémov, či už ide o vyhľadávanie

stípcov v projekte CAD alebo sémanticky blízkych obrázkov v databáze.

Odborníci pracujúci s databázami by mali venovať pozornosť vektorovým skladom. Ich rozšírenie naznačuje novú etapu vývoja databáz, v ktorej sa začínajú prelínati klasické relačné systémy a technológie orientované na umelú inteligenciu a vytvárajú hybridné riešenia budúcnosti.

Používatelia, ktorí vyvíjajú komplexné a rozsiahle aplikácie umelej inteligencie, budú používať špecializované databázy na vektorové vyhľadávanie. Zároveň tí, ktorí potrebujú len samostatné AI-funkcie na integráciu do existujúcich aplikácií, si s väčšou pravdepodobnosťou zvolia zabudované možnosti vektorového vyhľadávania v databázach, ktoré už používajú (PostgreSQL, Redis).

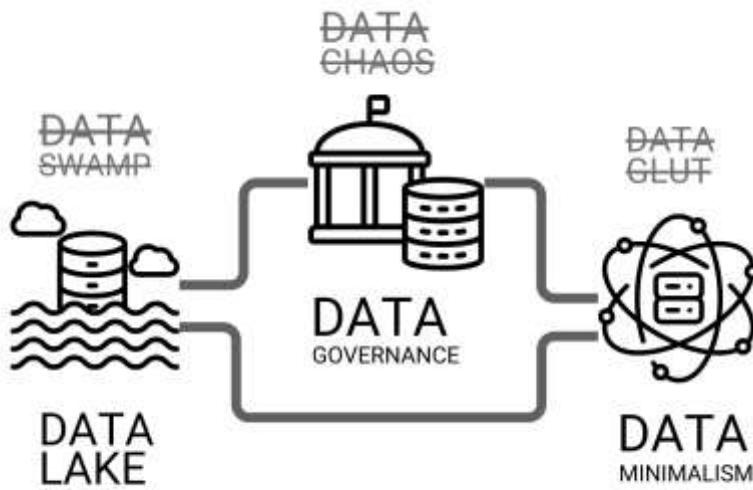
Hoci systémy ako DWH, Data Lake, CDE, PMIS, vektorové databázy a ďalšie ponúkajú rôzne prístupy k ukladaniu a správe údajov, ich účinnosť závisí nielen od ich architektúry, ale aj od toho, ako dobre sú samotné údaje organizované a spravované. Aj pri používaní moderných riešení - či už ide o vektorové databázy, klasické relačné DBMS alebo sklady typu Data Lake - môže nedostatok jasných pravidiel na správu, štruktúrovanie a aktualizáciu údajov viest k rovnakým problémom, s akými sa stretávajú používatelia pracujúci s rôznorodými súbormi a údajmi vo viacerých formátoch.

Bez správy údajov) sa aj tie najvýkonnejšie riešenia môžu stať chaotickými a neštruktúrovanými, čím sa z dátových jazier stanú dátové bažiny). Aby sa tomu spoločnosti vyhli, musia nielen zvoliť správnu architektúru úložiska, ale aj implementovať stratégie minimalizácie údajov), riadenia prístupu a kontroly kvality, aby sa z údajov stal účinný nástroj na rozhodovanie.

### **Riadenie údajov), minimalizmus údajov) a dátová bažina).**

Pochopenie a implementácia koncepcií správy údajov), minimalizácie údajov) a prevencie dátovej bažiny) sú klúčom k úspešnému riadeniu dátových skladov a dosahovaniu obchodnej hodnoty (obrázok 8.2-3).

Podľa štúdie spoločnosti Gartner (2017) 85% projektov veľkých dát zlyháva a jedným z klúčových dôvodov je nedostatočná kvalita údajov a ich správa [144].



Obr. 8.2-3 Medzi kľúčové aspekty správy údajov patria správa údajov a minimalizmus údajov.

Správa údajov (Data Governance) je základnou zložkou správy údajov, ktorá zabezpečuje, aby sa údaje používali vhodne a efektívne vo všetkých obchodných procesoch. Nejde len o stanovenie pravidiel a postupov, ale aj o zabezpečenie dostupnosti, spoľahlivosti a bezpečnosti údajov:

- Definovanie a klasifikácia údajov: jasné definovanie a klasifikácia entít umožňuje organizáciám pochopiť, aké entity sú v spoločnosti potrebné, a určiť, ako by sa mali používať.
- Prístupové práva a správa: vypracovanie politík a postupov pre prístup k údajom a ich správu zabezpečuje, že k určitým údajom majú prístup len oprávnení používateľia.
- Ochrana údajov pred vonkajšími hrozobami: Ochrana údajov pred vonkajšími hrozobami je kľúčovým aspektom správy údajov. Zahŕňa nielen technické opatrenia, ale aj školenie zamestnancov o základoch informačnej bezpečnosti.

Minimalizácia údajov (Data Minimalism) je prístup, ktorého cieľom je zredukovať údaje na najhodnotnejšie a najvýznamnejšie atribúty a entity v útvare (obrázok 8.2-4), a tým znížiť náklady a zlepšiť využitie údajov:

- Zjednodušenie rozhodovania: redukcia počtu objektov a ich atribútov na tie najdôležitejšie zjednodušuje rozhodovanie tým, že znižuje čas a zdroje potrebné na analýzu a spracovanie údajov.
- Zameranie sa na to, čo je dôležité: výber najdôležitejších entít a atribútov vám umožní zamierať sa na informácie, ktoré sú pre podnik skutočne dôležité, a eliminovať šum a nepotrebné údaje.
- Efektívne prideľovanie zdrojov: minimalizácia údajov umožňuje efektívnejšie prideľovanie zdrojov, znižuje náklady na ukladanie a spracovanie údajov, zlepšuje kvalitu a bezpečnosť údajov.

Logika práce s údajmi by nemala začínať ich samotným vytvorením (obr. 8.2-4), ale pochopením budúcich scenárov využitia týchto údajov ešte pred začatím procesu ich tvorby. Tento prístup umožňuje vopred definovať minimálne potrebné požiadavky na atribúty, ich typy a hraničné hodnoty. Tieto požiadavky tvoria základ pre vytvorenie správnych a stabilných entít v informačnom modeli.

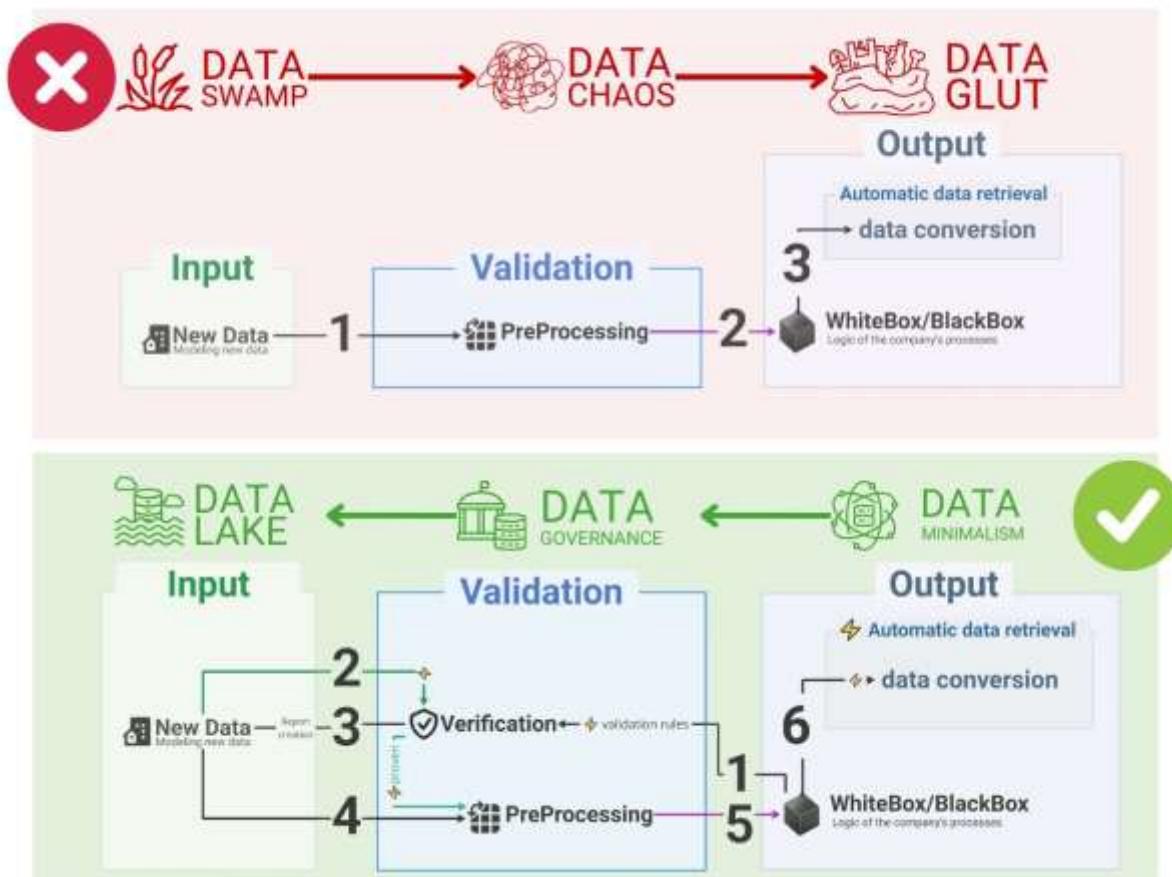
Predbežné pochopenie účelu a použitia údajov prispieva k vytvoreniu štruktúry vhodnej na analýzu. Podrobnejšie informácie o prístupoch k modelovaniu údajov na konceptuálnej, logickej a fyzickej úrovni boli uvedené v kapitole "Modelovanie údajov: konceptuálny, logický a fyzický model".

V tradičných obchodných procesoch stavebných firiem sa spracovanie údajov častejšie podobá vylievaniu údajov do bažiny, kde sa údaje najprv vytvoria a potom sa ich špecialisti snažia integrovať do iných systémov a nástrojov.

Dátová bažina (Data Swamp) je výsledkom nekontrolovaného zhromažďovania a ukladania údajov bez náležitej organizácie, štruktúrovania a riadenia, čo vedie k tomu, že údaje sú neštruktúrované, ťažko použiteľné a málo hodnotné.

Ako zabrániť tomu, aby sa tok informácií zmenil na bahisko:

- **Správa štruktúry údajov:** zabezpečenie štruktúrovaných a kategorizovaných údajov pomáha predchádzať zahľteniu údajmi tým, že ich robí prehľadnými a ľahko dostupnými.
- **Pochopenie a interpretácia** údajov: jasný opis pôvodu, úprav a významu údajov zabezpečuje ich správne pochopenie a interpretáciu.
- **Udržiavanie kvality údajov:** pravidelná údržba a čistenie údajov pomáha udržiavať kvalitu, relevantnosť a hodnotu údajov pre analytické a obchodné procesy.



Obrázok 8.2-4 Aby ste sa vyhli neporiadku v dátovom sklade, mali by ste proces vytvárania údajov začať zhromažďovaním požiadaviek na atribúty.

Integráciou zásad správy údajov a minimalizmu údajov do procesov správy údajov a aktívnym predchádzaním tomu, aby sa dátové sklady stali dátovými bažinami, môžu organizácie maximalizovať potenciál svojich údajov.

Ďalšou etapou vo vývoji práce s údajmi, po vyriešení otázok riadenia a minimalizmu, je štandardizácia automatického spracovania, zabezpečenie kvality a implementácia metód, ktoré umožňujú využitie údajov na analýzu, transformáciu a rozhodovanie. Tomu sa venujú metodiky DataOps a VectorOps, ktoré sa stávajú dôležitými nástrojmi pre firmy pracujúce s veľkými dátami a strojovým učením.

## DataOps a VectorOps: nové dátové štandardy

Zatiaľ čo Data Governance je zodpovedný za kontrolu a organizáciu údajov, DataOps pomáha zabezpečiť ich presnosť, konzistenciu a plynulý tok v rámci spoločnosti. To je obzvlášť dôležité pre viaceré obchodné prípady v stavebnictve, kde údaje vznikajú priebežne a vyžadujú si včasné spracovanie. Napríklad v situáciach, keď je potrebné synchronizovať informačné modely stavby, požiadavky na projekt a analytické správy medzi rôznymi systémami v priebehu jedného pracovného

dňa, môže byť úloha DataOps kľúčová. Umožňuje vytvoriť stabilné a opakovateľné procesy spracovania údajov, čím sa znižuje riziko oneskorenia a straty relevantnosti informácií.

Samotná správa údajov nestačí - je nevyhnutné, aby sa údaje nielen uchovávali, ale aj aktívne využívali v každodennej činnosti. Tu prichádza na rad DataOps - metodika zameraná na automatizáciu, integráciu a nepretržitý tok údajov.

**DataOps** sa zameriava na zlepšenie spolupráce, integrácie a automatizácie dátových tokov v organizáciách. Prijatie postupov DataOps podporuje presnosť, konzistenciu a dostupnosť údajov, čo je rozhodujúce pre aplikácie orientované na údaje.

Kľúčovými nástrojmi ekosystému DataOps sú Apache Airflow (obrázok 7.4-4) na orchestráciu pracovných tokov a Apache NiFi (obrázok 7.4-5) na smerovanie a transformáciu dátových tokov. Tieto technológie spoločne umožňujú flexibilné, spoľahlivé a škálovateľné dátové potrubia na automatické spracovanie, riadenie a integráciu informácií medzi systémami (viac informácií v kapitole "Automatický ETL -prevodník"). Pri implementácii prístupu DataOps do stavebných procesov je dôležité zohľadniť štyri základné aspekty:

1. **Ludia a nástroje sú dôležitejšie ako údaje:** oddelené úložiská údajov sa môžu považovať za hlavný problém, ale skutočnosť je zložitejšia. Okrem roztrieštenosti údajov zohráva významnú úlohu izolácia tímov a rozdielne nástroje, ktoré používajú. V stavebníctve pracujú s údajmi špecialisti z rôznych odborov: dátoví inžinieri a analytici, tímy BI a vizualizácie, ako aj odborníci na riadenie projektov a kvalitu. Každý z nich má odlišné spôsoby práce, preto sa stáva dôležitým vytvoriť ekosystém, v ktorom údaje voľne prúdia medzi účastníkmi a poskytujú jednotnú, konzistentnú verziu informácií.
2. **Automatizácia testovania a zisťovania chýb:** Údaje o stavbe vždy obsahujú chyby, či už ide o nepresnosti v modeloch, chyby vo výpočtoch alebo neaktuálne špecifikácie. Pravidelné testovanie údajov a odstraňovanie opakujúcich sa chýb môže výrazne zlepšiť kvalitu údajov. V rámci DataOps je potrebné zaviesť automatizované kontrolné a validačné mechanizmy, ktoré monitorujú správnosť údajov, analyzujú chyby a identifikujú vzory a zachytávajú a riešia systémové chyby v každom pracovnom postupe. Čím vyšší je stupeň automatizovanej validácie, tým vyššia je celková kvalita údajov a tým nižšia je pravdepodobnosť výskytu chýb v konečných fázach.
3. **Údaje by sa mali testovať rovnako ako programový kód:** väčšina stavebných aplikácií je založená na spracovaní údajov, ale ich kontrola sa často ponecháva na sekundárne úlohy. Ak sa modely strojového učenia trénujú na nepresných údajoch, vedie to k nesprávnym predpovediam a finančným stratám. V rámci DataOps by sa údaje mali podrobiť rovnako prísnej kontrole ako programový kód: logické kontroly, záťažové testy, vyhodnotenie správania modelov pri zmene vstupných hodnôt. Len overené a spoľahlivé údaje sa môžu používať ako základ pre rozhodnutia manažmentu.
4. **Pozorovateľnosť údajov bez ohrozenia výkonnosti:** monitorovanie údajov nie je len súbor metrík, ale strategický nástroj riadenia kvality. Aby DataOps fungoval efektívne, pozorovateľnosť musí byť zabudovaná vo všetkých fázach spracovania údajov, od návrhu až po prevádzku. Zároveň je dôležité, aby monitorovanie nespomaľovalo systém. V stavebných projektoch je veľmi dôležité nielen zbierať údaje, ale robiť to takým spôsobom, aby sa nijako

nenarušila práca odborníkov (napr. projektantov), ktorí údaje vytvárajú. Táto rovnováha umožňuje kontrolovať kvalitu údajov bez toho, aby bola ohrozená produktivita.

DataOps nie je pre dátových vedcov ďalšou záťažou, ale základom ich práce. Implementáciou DataOps môžu stavebné spoločnosti prejsť od chaotickej správy dát k efektívному ekosystému, v ktorom dátu pracujú pre podnikanie.

VectorOps zasa predstavuje ďalšiu etapu vývoja DataOps, zameranú na spracovanie, ukladanie a analýzu viacozmerných vektorových údajov (o ktorých sa hovorilo v predchádzajúcej kapitole). Je to dôležité najmä v oblastiach, ako sú digitálne dvojčatá, modely neurónových sietí a sémantické vyhľadávanie, ktoré začínajú prichádzať do stavebníctva. VectorOps sa spolieha na vektorové databázy na efektívne ukladanie, indexovanie a vyhľadávanie viacozmerných reprezentácií objektov.

**VectorOps** je ďalším krokom po DataOps, zameraným na spracovanie, analýzu a využívanie vektorových údajov v stavebníctve. Na rozdiel od DataOps, ktorý sa zameriava na tok, konzistenciu a kvalitu údajov, VectorOps sa zameriava na správu viacozmerných reprezentácií objektov potrebných na strojové učenie.

Na rozdiel od tradičných prístupov umožňuje VectorOps dosiahnuť presnejší popis objektov, čo je rozhodujúce pre digitálne dvojčatá, systémy generatívneho navrhovania a automatickú detekciu chýb v údajoch CAD konvertovaných do vektorového formátu. Kombinovaná implementácia DataOps a VectorOps tvorí pevný základ pre škálovateľnú, automatizovanú prácu s veľkými objemami informácií - od klasických tabuľiek až po sémanticky bohaté priestorové modely.

## Ďalšie kroky: od chaotického ukladania k štruktúrovanému ukladaniu

Tradičné prístupy k budovaniu dátových úložísk často vedú k vytváraniu rôznorodých "informačných síl", v ktorých sú dôležité poznatky nedostupné pre analýzu a rozhodovanie. Moderné koncepcie ukladania dát, ako napríklad dátový sklad, dátové jazero a ich hybridy, umožňujú zjednotiť rôznorodé informácie a sprístupniť ich centralizovaným spôsobom pre dátové toky a business intelligence. Je dôležité nielen vybrať správnu architektúru úložiska, ale aj zaviesť správu údajov) a minimalizmus údajov), aby sa z úložísk nestali nekontrolovatelné dátové bažiny).

Na zhrnutie tejto časti je vhodné zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať preberané koncepty na vaše každodenné úlohy:

- Výber efektívnych formátov ukladania údajov
  - Prechod z formátov CSV a XLSX na efektívnejšie formáty (Apache Parquet, ORC) na ukladanie veľkého množstva údajov
  - Implementácia systému verzií údajov na sledovanie zmien
  - Používanie metadát na opis štruktúry a pôvodu informácií
- Vytvorenie jednotnej architektúry podnikových údajov
  - Porovnanie rôznych architektúr ukladania: RDBMS, DWH a Data Lake. Vyberte si tú, ktorá

najlepšie vyhovuje vašim potrebám škálovateľnosti, integrácie zdrojov a analytického spracovania.

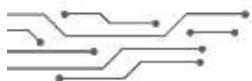
- Navrhnite mapu procesov na extrakciu, načítanie a transformáciu údajov (ETL) z rôznych zdrojov pre svoje úlohy. Na vizualizáciu kľúčových krokov a integračných bodov použite vizualizačné nástroje, napríklad Miro, Lucidchart alebo Draw.io

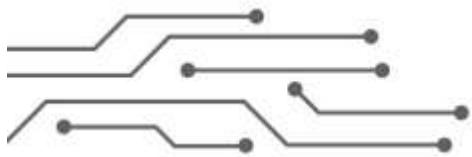
#### ■ Implementácia postupov správy údajov a minimalizácia údajov

- Dodržiavajte prístup minimalizmu údajov - ukladajte a spracúvajte len to, čo je skutočne cenné.
- Implementácia zásad správy údajov - definovanie zodpovednosti za údaje, zabezpečenie kvality a transparentnosti
- Ďalšie informácie o zásadách správy údajov a konceptoch DataOps, VectorOps
- Definovanie kritérií kvality údajov a postupov validácie údajov v rámci DataOps

Dobre organizované ukladanie údajov vytvára základ pre centralizáciu analytických procesov spoločnosti. Prechod od chaotického hromadenia súborov k štruktúrovaným úložiskám umožňuje premeniť informácie na strategické aktívum, ktoré pomáha prijímať informované rozhodnutia a zlepšovať efektívnosť podnikových procesov.

Po automatizácii a štandardizácii procesov zberu, transformácie, analýzy a štruktúrovaného ukladania údajov je ďalšou etapou digitálnej transformácie plnohodnotné spracovanie veľkých objemov údajov.





## **IX ČASŤ**

# **VEĽKÉ DÁTA, STROJOVÉ UČENIE A PREDPOVEDE**

Deviata časť sa zameriava na veľké dátá, strojové učenie a prediktívnu analýzu v stavebníctve. Skúma prechod od intuitívneho rozhodovania k objektívnym analýzam na základe historických údajov. Na praktických príkladoch sa demonštruje analýza veľkých dát v stavebníctve - od rozboru súboru údajov o stavebných povoleniach v San Franciscu až po spracovanie CAD - projektov s miliónmi prvkov. Osobitná pozornosť sa venuje metódam strojového učenia na predpovedanie nákladov a harmonogramu stavebných projektov, pričom sa podrobne rozoberá lineárna regresia a algoritmy k-najbližších susedov. Ukazuje sa, ako sa štruktúrované údaje stávajú základom prediktívnych modelov na posúdenie rizík, optimalizáciu zdrojov a zlepšenie efektívnosti riadenia projektov. V tejto časti sa uvádzajú aj odporúčania na výber reprezentatívnych vzoriek údajov a vysvetľuje sa, prečo na efektívne analýzy nie sú vždy potrebné veľké súbory údajov.

## KAPITOLA 9.1.

# VEĽKÉ OBJEMY ÚDAJOV A ICH ANALÝZA

### **Veľké dáta v stavebnictve: od intuície k predvídateľnosti**

Pojem "big data" nemá striktnú definíciu. Tento pojem sa pôvodne objavil, keď objem informácií začal presahovať možnosti tradičných metód ich spracovania. Dnes sa objem a zložitosť údajov v mnohých odvetviach vrátane stavebnictva zvýšili natoľko, že sa nezmestia do lokálnej pamäte počítačov a na ich spracovanie je potrebné použiť nové technológie.

Podstatou práce s veľkými dátami nie je len ukladanie a spracovanie, ale aj prediktívne schopnosti. V stavebnictve otvárajú Big Data cestu od intuitívnych rozhodnutí založených na subjektívnej interpretácii tabuľiek a vizualizácií (ako už bolo spomenuté) k informovaným prognózam podloženým reálnymi pozorovaniami a štatistikami.

Na rozdiel od všeobecného presvedčenia nie je cieľom práce s veľkými dátami "prinútiť stroj myslieť ako človek", ale použiť matematické modely a algoritmy na analýzu obrovského množstva údajov s cieľom identifikovať vzory, predpovedať udalosti a optimalizovať procesy.

**Big Data** nie sú chladným svetom algoritmov bez ľudského vplyvu. Naopak, veľké dáta fungujú v spojení s našimi inštinktmi, chybami a kreativitou. Práve nedokonalosť ľudského mysenia nám umožňuje nachádzať neštandardné riešenia a robiť prelomové objavy.

S rozvojom digitálnych technológií začal stavebný priemysel aktívne využívať techniky spracovania údajov, ktoré pochádzajú z odvetvia informačných technológií. Vďaka nástrojom, ako sú Pandas a Apache Parquet, možno kombinovať štruktúrované a neštruktúrované údaje, čím sa zjednodušuje prístup k informáciám a znižuje sa strata pri analýze, zatiaľ čo veľké súbory údajov z dokumentov alebo projektov CAD (obr. 9.2-10 - obr. 9.2-12) umožňujú zhromažďovať, analyzovať a predvídať údaje vo všetkých fázach životného cyklu projektu.

Big Data majú transformačný vplyv na stavebnictvo a potenciálne ho ovplyvňujú rôznymi spôsobmi. Aplikácia technológií Big Data prináša výsledky vo viacerých klúčových oblastiach, vrátane napr:

- **Analýza investičného potenciálu** - prognóza ziskovosti a doby návratnosti projektov na základe údajov z predchádzajúcich zariadení.
- **Prediktívna údržba** - identifikácia pravdepodobných porúch zariadenia skôr, ako k nim skutočne dôjde, čo znižuje prestoje.
- **Optimalizácia dodávateľského reťazca** - predvídanie porúch a zvyšovanie efektívnosti logistiky.
- **Analýzy energetickej účinnosti** - pomoc pri navrhovaní nízkoenergetických budov.
- **Monitorovanie bezpečnosti** - používanie senzorov a nositeľných zariadení na monitorovanie podmienok na stavenisku.

- **Kontrola kvality** - monitorovanie dodržiavania procesných nariem v reálnom čase.
- **Riadenie pracovnej sily** - analýza výkonnosti a prognóza personálnych požiadaviek.

Ťažko nájsť oblasť v stavebníctve, v ktorej by nebola žiadana analýza údajov a predpovede. Hlavnou výhodou predikčných algoritmov je ich schopnosť samoučiť sa a neustále sa zlepšovať podľa toho, ako sa hromadia údaje.

V blízkej budúcnosti bude umelá inteligencia nielen pomáhať stavitelom, ale bude robiť kľúčové rozhodnutia - od procesov navrhovania až po otázky prevádzky budov.

Viac o tom, ako sa vytvárajú predpovede a používajú modely učenia, sa dozviete v ďalšej časti knihy "Strojové učenie a predpovede".

Prechod na plnohodnotnú prácu s veľkými dátami si vyžaduje zmenu samotného prístupu k analytike. Kým klasické systémy, ktoré sme doteraz posudzovali, sa zameriavalí na vzťahy príčiny a následku, analytika veľkých dát sa presúva na hľadanie štatistických vzorcov a korelácií, ktoré umožňujú identifikovať skryté vzťahy a predpovedať správanie objektov aj bez úplného pochopenia všetkých faktorov.

## **Spochybňovanie uskutočnitelnosti veľkých dát: korelácia, štatistika a výber vzoriek údajov**

Tradične sa konštrukcia zakladala na subjektívnych hypotézach a osobných skúsenostiah. Inžinieri predpokladali - s určitou mierou pravdepodobnosti - ako sa bude materiál správať, aké zaťaženie konštrukcia vydrží a ako dlho bude projekt trvať. Tieto predpoklady sa overovali v praxi, často na úkor času, zdrojov a budúcich rizík.

S príchodom veľkých dát sa tento prístup dramaticky mení: rozhodnutia sa už nerobia na základe intuitívnych predstáv, ale ako výsledok analýzy rozsiahlych súborov dát. Výstavba postupne prestáva byť umením intuúcie a stáva sa presnou vedou o predpovediach.

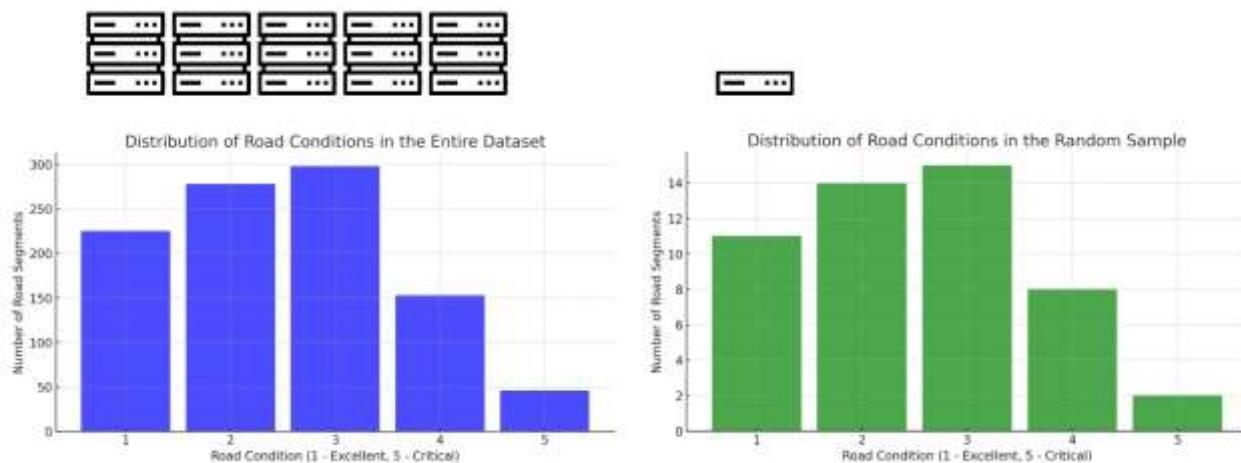
Prechod k myšlienke využívania veľkých dát nevyhnutne vyvoláva dôležitú otázku: aké rozhodujúce je množstvo dát a koľko informácií je skutočne potrebných na spoľahlivú prediktívnu analýzu? Všeobecne rozšírené presvedčenie, že "čím viac údajov, tým vyššia presnosť", sa v praxi nie vždy ukazuje ako štatisticky platné.

Už v roku 1934 štatistik Jerzy Neumann dokázal [145], že kľúč k presnosti štatistických záverov nespočíva ani tak v množstve údajov, ako v ich reprezentatívnosti a náhodnosti výberu.

Platí to najmä v stavebníctve, kde sa zhromažďuje veľké množstvo údajov pomocou senzorov internetu vecí, skenerov, monitorovacích kamier, dronov a dokonca aj viacformátových modelov CAD, čo zvyšuje riziko mŕtvykh bodov, odľahlých hodnôt a skreslenia údajov.

Uveďme si príklad monitorovania stavu povrchu ciest. Kompletný súbor údajov o všetkých úsekoch

ciest môže mať veľkosť X GB a jeho spracovanie môže trvať približne jeden deň. Zároveň náhodná vzorka zahŕňajúca len každý 50. úsek cesty by zabrala len X/50 GB a spracovala by sa za pol hodiny, pričom by poskytovala podobnú presnosť odhadov pre určité výpočty (obrázok 9.1-1).



Obrázok 9.1-1 Histogramy stavu vozovky: úplný súbor údajov a náhodný výber vzorky vykazujú rovnaké výsledky.

Klúčom k úspešnej analýze údajov teda často nemusí byť množstvo údajov, ale reprezentatívnosť vzorky a kvalita použitých metód spracovania. Prechod na náhodný výber vzoriek a selektívnejší prístup si vyžaduje zmenu myslenia v stavebnictve. V minulosti sa spoločnosti riadili logikou "čím viac údajov, tým lepšie" v presvedčení, že pokrytie všetkých možných ukazovateľov maximalizuje presnosť.

Tento prístup pripomína populárny omyl z projektového manažmentu: "čím viac odborníkov prilákam, tým bude práca efektívnejšia". Podobne ako pri ľudských zdrojoch sú však dôležitejšie kvalita a nástroje než kvantita. Bez zohľadnenia vzájomných vzťahov (korelácií) medzi údajmi alebo účastníkmi projektu môže zvyšovanie objemu viesť len k šumu, skresleniu, duplicitne a zbytočnému plynaniu.

Nakoniec sa často ukáže, že je oveľa produktívnejšie mať k dispozícii menší, ale kvalitatívne pripravený súbor údajov, ktorý dokáže vytvoriť stabilné a primerané prognózy, ako sa spoliehať na obrovské, ale chaotické informácie obsahujúce množstvo protichodných signálov.

Nadmerný objem údajov nielenže nezaručuje väčšiu presnosť, ale môže viesť aj k skresleným záverom v dôsledku prítomnosti šumu, nadbytočných prvkov, skrytých korelácií a irelevantných informácií. Za takýchto okolností sa zvyšuje riziko nadmerného prispôsobenia modelov a znižuje sa spoľahlivosť analytických výsledkov.

V stavebnictve je hlavnou výzvou pri práci s veľkými údajmi určenie optimálneho množstva a kvality údajov. Napríklad pri monitorovaní stavu betónových konštrukcií môže používanie tisícov snímačov a zbieranie informácií každú minútu zahľubiť systém ukladania a analýzy. Ak však vykonáte korelačnú analýzu a vyberiete 10% najinformatívnejších snímačov, môžete získať takmer identickú presnosť predpovedí, pričom vynaložíte mnohonásobne, niekedy desiatky a stovky krát menej prostriedkov.

Použitie menšej podmnožiny údajov znižuje objem potrebného úložiska aj čas spracovania, čo výrazne znižuje náklady na ukladanie a analýzu údajov a často robí z náhodného výberu vzoriek ideálne riešenie pre prediktívnu analýzu, najmä v prípade veľkých infraštruktúrnych projektov alebo pri práci v reálnom čase. V konečnom dôsledku o efektívnosti stavebných procesov nerozhoduje množstvo zozbieraných údajov, ale kvalita ich analýzy. Bez kritického prístupu a dôkladnej analýzy môžu údaje viesť k nesprávnym záverom.

Po určitom množstve údajov prináša každá nová jednotka informácií čoraz menej užitočných výsledkov. Namiesto nekonečného zhromažďovania informácií je dôležité zameriť sa na ich reprezentatívnosť a metódy analýzy (obrázok 9.2-2).

Tento jav dobre opísal Allen Wallis [146], ktorý ilustruje použitie štatistických metód na príklade testovania dvoch alternatívnych konštrukcií projektilov amerického námorníctva.

*Námorníctvo testovalo dva alternatívne návrhy projektilov (A a B) sériou párových výstrelov. V každom kole A dostane 1 alebo 0 podľa toho, či je jeho výkon lepší alebo horší ako výkon B, a naopak. Štandardný štatistický prístup zahŕňa vykonanie pevného počtu pokusov (napr. 1 000) a určenie víťaza na základe percentuálneho rozdelenia (napr. ak A získa 1 vo viac ako 53% prípadov, považuje sa za najlepší). Keď Allen Wallis diskutoval o takomto probléme s kapitánom (námorníctva) Garrettom L. Schuylerom, kapitán namietol, že takýto test, povedané Allenovou rečou, môže byť zbytočný. Keby bol na mieste múdry a skúsený ordonančný dôstojník, akým bol Schuyler, po prvých niekoľkých stovkách [výstrelov] by videl, že experiment netreba ukončiť, bud' preto, že nová metóda je zjavne horšia, alebo preto, že je zjavne lepšia, ako sa dúfalo [146].*

- Štatistická výskumná skupina vlády USA na Kolumbijskej univerzite, obdobie druhej svetovej vojny

Tento princíp sa široko používa v rôznych priemyselných odvetviach. Napríklad v medicíne sa klinické skúšky nových liekov vykonávajú na náhodných vzorkách pacientov, čo umožňuje získať štatisticky významné výsledky bez potreby testovať liek na celej populácii ľudí žijúcich na našej planéte. V ekonómii a sociológii sa vykonávajú reprezentatívne prieskumy, ktoré odrážajú názory spoločnosti bez toho, aby bolo potrebné vypočuť každého obyvateľa krajiny.

Podobne ako vlády a výskumné organizácie vykonávajú prieskum v malých populáciách, aby pochopili všeobecné sociálne trendy, aj spoločnosti v stavebnictve môžu využívať náhodné vzorky údajov na efektívne monitorovanie a vytváranie prognóz pre riadenie projektov (obrázok 9.1-1).

*Big data môžu zmeniť prístup k sociálnym vedám, ale nenahradia štatistický zdravý rozum [147].*

- Thomas Landsall-Welfair, "Predpovedanie aktuálnej nálady národa", Significance v. 9(4), 2012 .

Z hľadiska úspory zdrojov je pri zhromažďovaní údajov na účely budúcich predpovedí a rozhodovania dôležité odpovedať na otázku: má zmysel vynakladať značné zdroje na zhromažďovanie a spracovanie obrovských súborov údajov, keď možno použiť oveľa menší a lacnejší súbor testovacích údajov, ktorý možno postupne zväčšovať? Účinnosť náhodného výberu vzoriek ukazuje, že spoločnosti môžu znížiť náklady na zber a školenie modelov desiatky alebo dokonca tisícky krát výberom metód zberu údajov, ktoré nevyžadujú komplexné pokrytie, ale stále poskytujú dostatočnú presnosť a reprezentatívnosť. Tento prístup umožňuje aj malým spoločnostiam dosahovať výsledky na úrovni veľkých korporácií s použitím výrazne menších zdrojov a objemov údajov, čo je dôležité pre spoločnosti, ktoré chcú optimalizovať náklady a urýchliť informované rozhodovanie s využitím malých zdrojov. V nasledujúcich kapitolách preskúmajte príklady analýzy a prediktívnej analýzy založenej na verejných súboroch údajov s využitím nástrojov big data.

## **Veľké dáta: analýza údajov zo súboru miliónov stavebných povolení v San Franciscu**

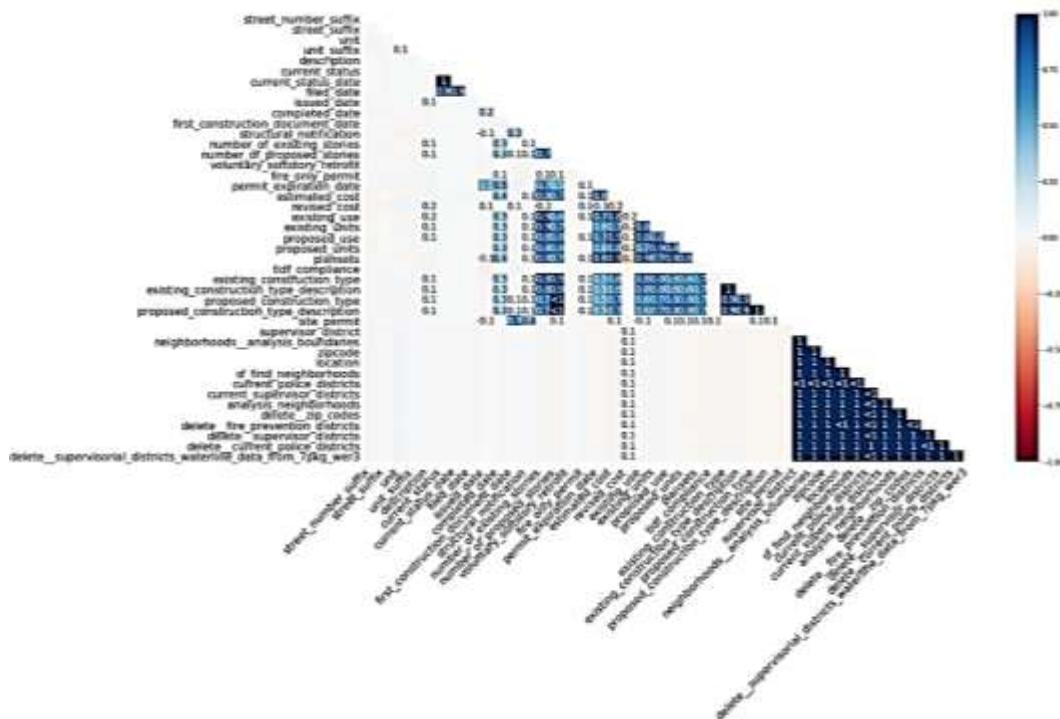
Práca s otvorenými súbormi údajov poskytuje jedinečnú príležitosť uplatniť v praxi zásady uvedené v predchádzajúcich kapitolách: uvážlivý výber prvkov, reprezentatívny výber vzoriek, vizualizácia a kritická analýza. V tejto kapitole budeme skúmať, ako možno pomocou otvorených údajov skúmať zložité javy, ako je napríklad stavebná činnosť vo veľkom meste - konkrétnie viac ako milión záznamov o stavebných povoleniach v San Franciscu

Verejne dostupné údaje o viac ako jednom milióne stavebných povolení (obrázok 9.1-2) (záznamy v dvoch súboroch údajov vo formáte CSV) od "San Francisco Department of Buildings" [148] nám umožňujú použiť nespracované CSV -tabuľky nielen na analýzu stavebnej činnosti v meste, ale aj na kritickú analýzu nedávnych trendov a histórie stavebníctva v San Franciscu za posledných 40 rokov, od roku 1980 do roku 2019.

Príklady kódu použité na vytvorenie vizualizácie súboru údajov (obrázky 9.1-3- Obrázky 9.1-8), ako aj vizuálne grafy s kódom, vysvetleniami a komentárimi nájdete na platforme Kaggle vyhľadaním "San Francisco. Sektor stavebníctva 1980-2019". [149].

count 1.137695e+06		Building Permits on or after January 1, 2013			Building Permits before January 1, 2013		
permit_creation_date	description	current_status	current_status_date	filed_date	issued_date	completed_date	
07/01/1998	repair stucco	complete		07/07/1998	07/01/1998	07/07/1998	
12/13/2004	reroofing	expired		01/24/2006	12/13/2004	12/13/2004	
02/18/1992	install auto fire spks.	complete		06/29/1992	02/18/1992	03/18/1992	
permit_number	permit_expiration_date	estimated_cost	revised_cost	existing_use	Zipcode	Location	
362780	9812394	11/01/1998	780.0	NaN	1 family dwelling	94123.0 (37.796246876498, -122.4322541443574)	
570017	200412131233	06/13/2005	9000.0	9000.0	apartments	94127.0 (37.729258516008388, -122.4644245957462)	
198411	9202396	09/18/1992	9000.0	NaN	apartments	94111.0 (37.795086002552974, -122.39593224461805)	

Obr. 9.1-2 Datasetsy obsahujú informácie o vydaných stavebných povoleniach s rôznymi atribútmi objektov.

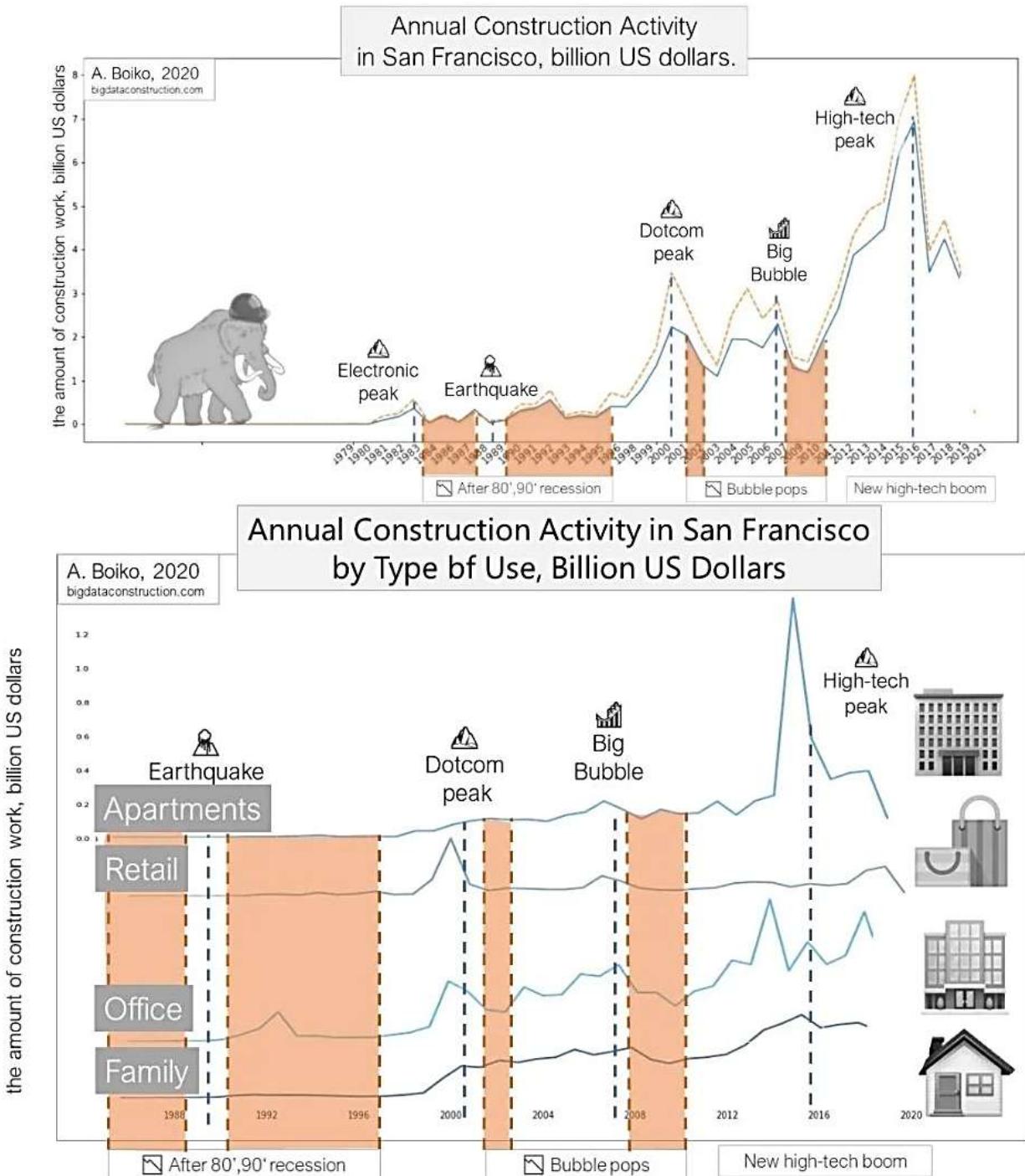


Obr. 9.1-3 Tepelná mapa (Pandas a Seaborn), ktorá vizualizuje všetky atribúty súboru údajov a pomáha identifikovať vzťahy medzi dvojicami atribútov.

Z tabuľky, ktorú poskytlo oddelenie budov v San Franciscu (obrázok 9.1-2), nie sú zrejmé žiadne trendy ani závery. Suché čísla v tabuľkovej forme nie sú podkladom pre rozhodovanie. Aby boli údaje vizuálne zrozumiteľné, ako sa podrobne uvádzajú v kapitolách o vizualizácii údajov, mali by sa vizualizovať pomocou rôznych knižníc, o ktorých sa hovorí v siedmej časti knihy na tému "ETL a vizualizácia výsledkov vo forme grafov".

Analýzou údajov pomocou knižníc Pandas DataFrame a Python visualisation libraries, o hodnote 1 137 695 povolení [148], môžeme dospiť k záveru, že stavebná aktivita v San Franciscu je úzko spojená s ekonomickými cyklami, najmä v prosperujúcom technologickom priemysle Silicon Valley (obrázok 9.1-4).

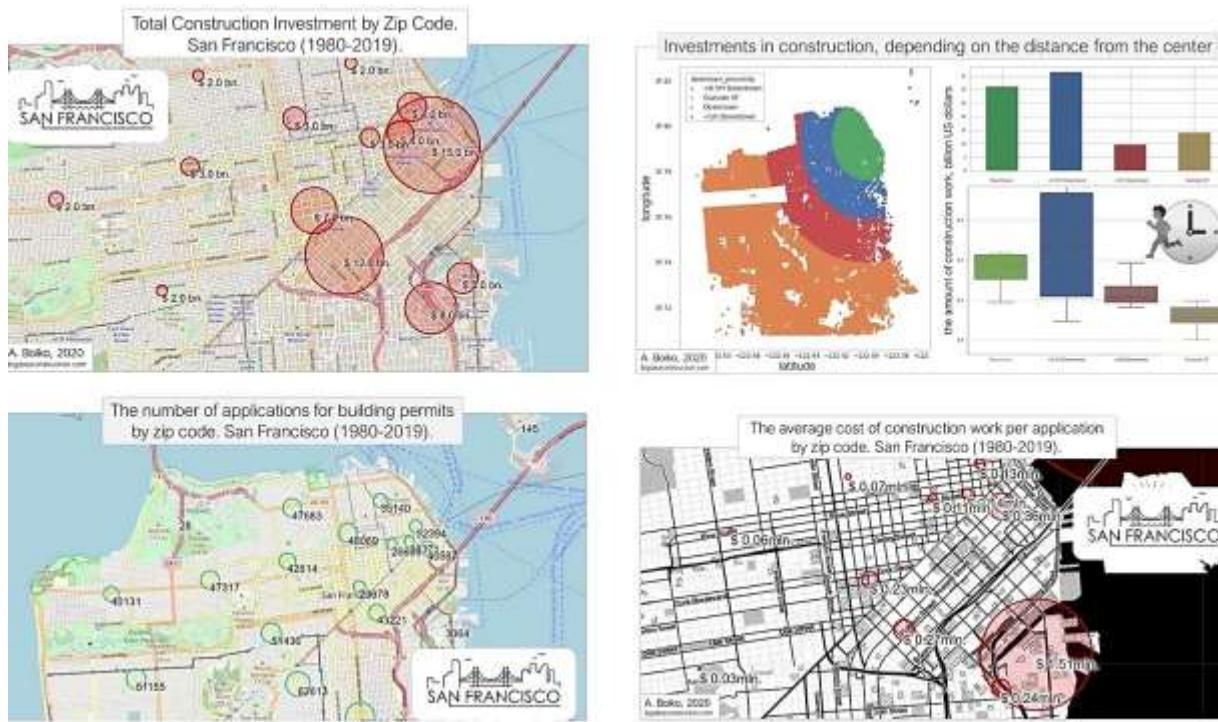
Hospodárske konjunktúry a krízy majú významný vplyv na počet a hodnotu stavebných projektov. Napríklad prvý vrchol stavebnej aktivity sa zhodoval s boomom elektroniky v polovici 80. rokov (použité Pandas a Matplotlib) a následné vrcholy a poklesy súviseli s bublinou dot-com a technologickým boomom v posledných rokoch.



Obr. 9.1-4 V sektore nehnuteľností v San Franciscu investície korelujú s technologickým rozvojom Silicon Valley.

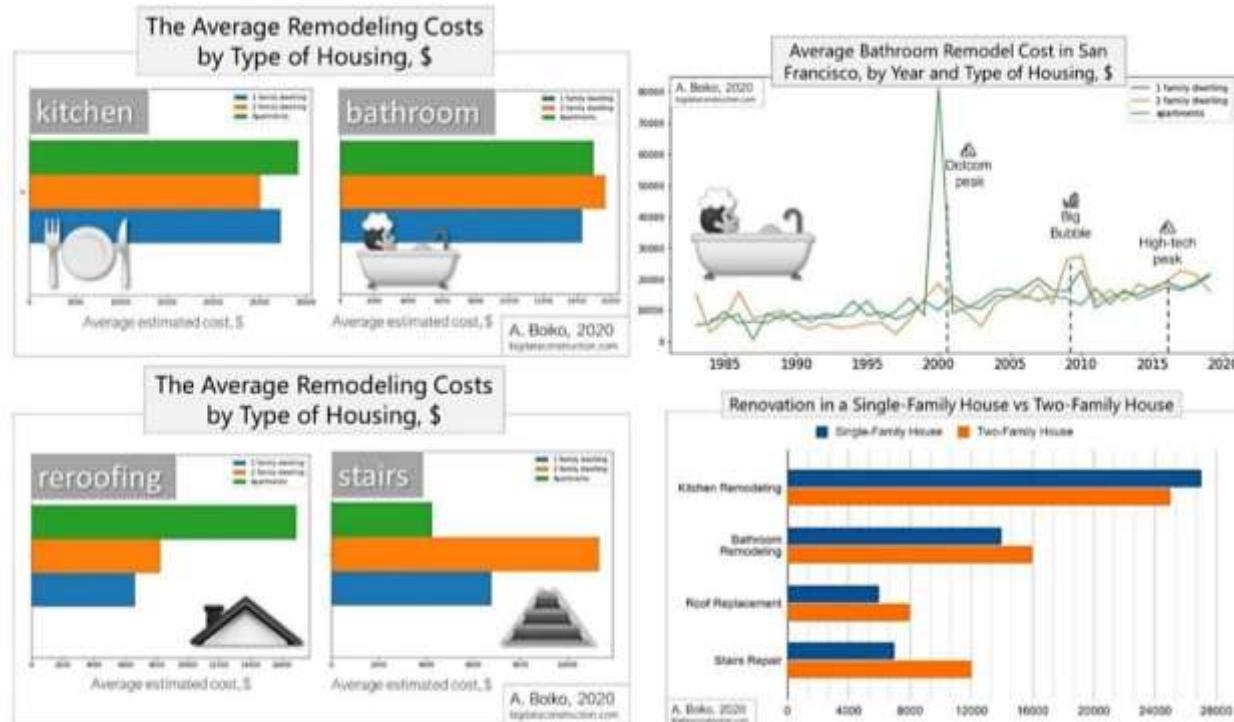
Analýza údajov naznačuje, že v San Franciscu sa väčšina 91,5 miliárd dolárov investovaných do výstavby a rekonštrukcie za posledné desaťročie - takmer 75% - sústredzuje v centre mesta (obrázok 9.1-5 - použité vizualizačné knižnice Pandas a Folium) a v okruhu 2 km od centra mesta, čo odráža vyššiu hustotu investícií v týchto centrálnych zónach.

Priemerné náklady na stavebné povolenia sa v jednotlivých štvrtiach výrazne líšia, pričom žiadosti v centre mesta sú trikrát drahšie ako žiadosti mimo centra mesta, a to z dôvodu vyšších nákladov na pozemky, prácu, materiály a prísnych stavebných predpisov, ktoré vyžadujú drahšie materiály na zlepšenie energetickej účinnosti.



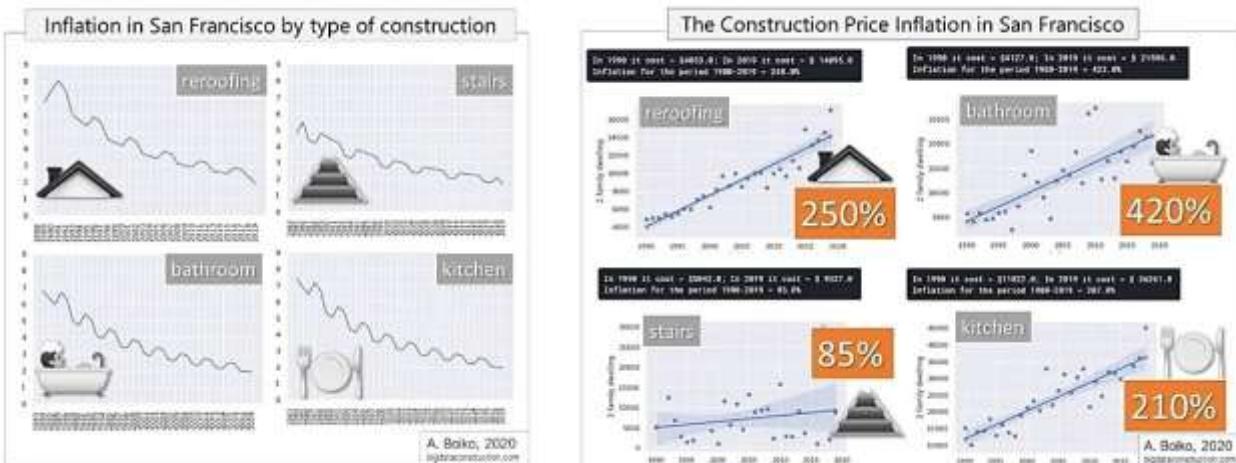
Obrázok 9.1-5 V San Franciscu je 75% stavebných investícií (91,5 miliardy USD) sústredených v centre mesta.

Súbor údajov tiež umožňuje výpočet priemerných cien opráv nielen podľa typu domu, ale aj podľa mestských štvrtí a jednotlivých adres (poštových smerovacích čísel). V San Franciscu dynamika nákladov na opravy domov vykazuje odlišné trendy pre rôzne typy rekonštrukcií a bývania (obrázok 9.1-6 - použité Pandas a Matplotlib). Renovácia kuchyne je výrazne drahšia ako renovácia kúpeľne: priemerná renovácia kuchyne v jednorodinnom dome stojí približne 28 000 USD v porovnaní s 25 000 USD v dvojrodinnom dome.



Obrázok 9.1-6 V SF stojí renovácia kuchyne takmer dvakrát viac ako renovácia kúpeľne a majitelia domov si musia každý mesiac odkladáť \$ 350 počas 15 rokov, aby pokryli náklady na väčšie opravy domu.

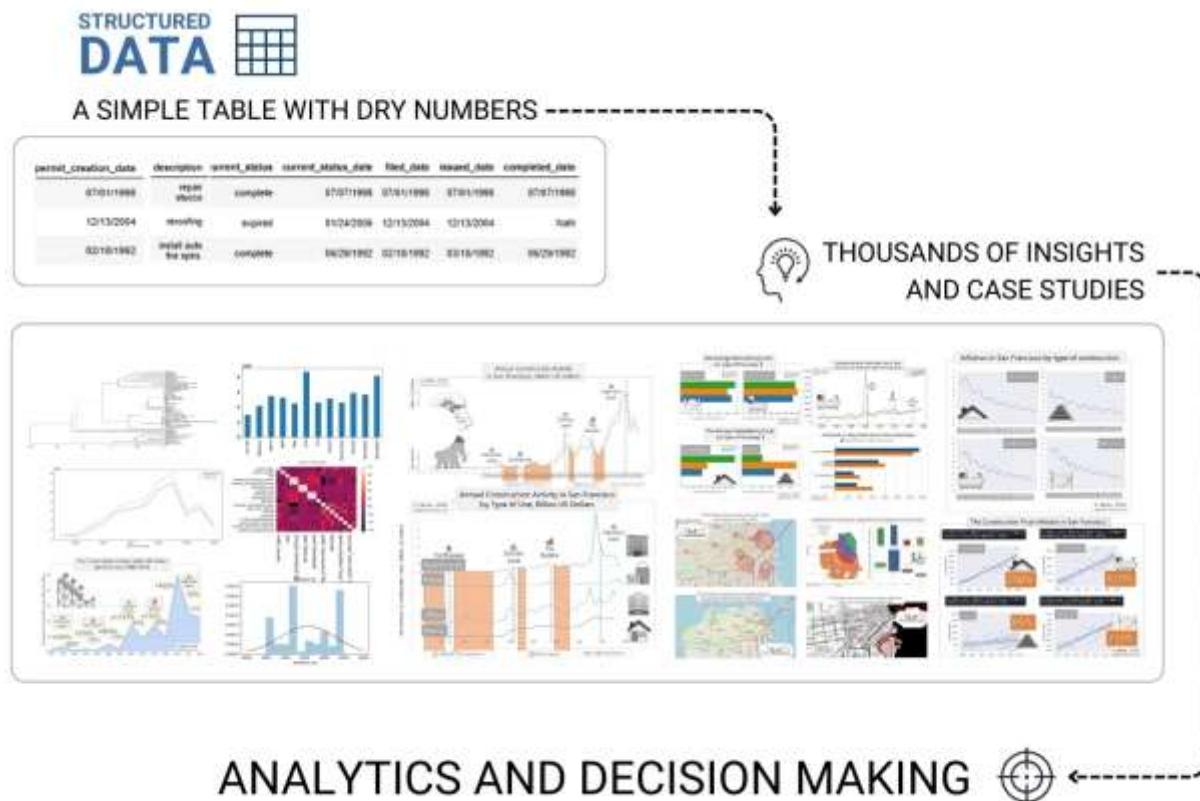
Infláciu stavebných nákladov v San Franciscu v priebehu rokov možno sledovať na základe analýzy údajov zoskupených podľa typu bývania a roku (obrázok 9.1-7 - použitý Pandas a Seaborn), z ktorého vyplýva neustály nárast priemerných nákladov na opravy od roku 1990 a odhaluje krátkodobé trojročné cykly v nákladoch na opravy viacbytových budov.



Obrázok 9.1-7 Od roku 1980 do roku 2019 sa náklady na rekonštrukciu kúpeľne v SF zvýšili päťnásobne, zatiaľ čo cena rekonštrukcie strechy a kuchyne sa strojnásobil a rekonštrukcia schodiska sa zvýšila len o 85%.

Štúdia verejných údajov zo sanfranciského stavebného úradu (obrázok 9.1-3) ukazuje, že stavebné náklady v meste sú veľmi premenlivé a často nepredvídateľné, ovplyvnené rôznymi faktormi. Medzi tieto faktory patrí hospodársky rast, technologické inovácie a jedinečné požiadavky rôznych typov bývania.

V minulosti si takéto analýzy vyžadovali hlboké znalosti programovania a analytiky. S príchodom nástrojov LLM- sa však tento proces stal prístupným a zrozumiteľným pre široké spektrum odborníkov v stavebnictve, od inžinierov v projekčných oddeleniach až po vrcholový manažment.



Obr. 9.1-8 Prechod na vizuálne zrozumiteľné údaje umožňuje automatizované rozhodovanie prostredníctvom rozpoznávania skrytých vzorcov.

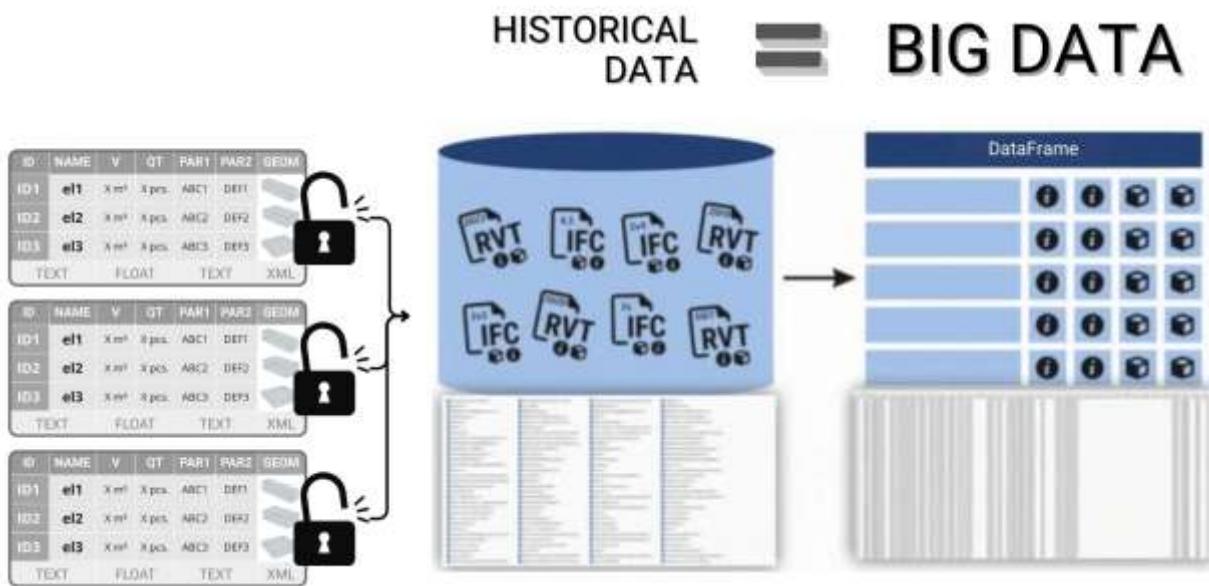
Tak ako sme analyzovali údaje z tabuľkového súboru údajov "San Francisco Building Authority", môžeme vizualizovať a analyzoovať akýkoľvek súbor údajov - od obrázkov a dokumentov až po údaje internetu vecí, alebo údaje z odvodených databáz CAD.

### Príklad veľkých dát na základe údajov CAD (BIM)

V nasledujúcom príklade budeme analyzovať veľký súbor údajov s použitím údajov z rôznych nástrojov CAD (BIM). Na zber a vytvorenie veľkého súboru údajov sa použil špecializovaný automatizovaný webový prehľadávač (skript) nakonfigurovaný na automatické vyhľadávanie a zber návrhových súborov z webových stránok ponúkajúcich bezplatné architektonické modely vo formátoch RVT a IFC.

Za niekoľko dní crawler úspešne našiel a stiahol 4 596 súborov IFC a 6 471 súborov RVT a 156 024 súborov DWG[149].

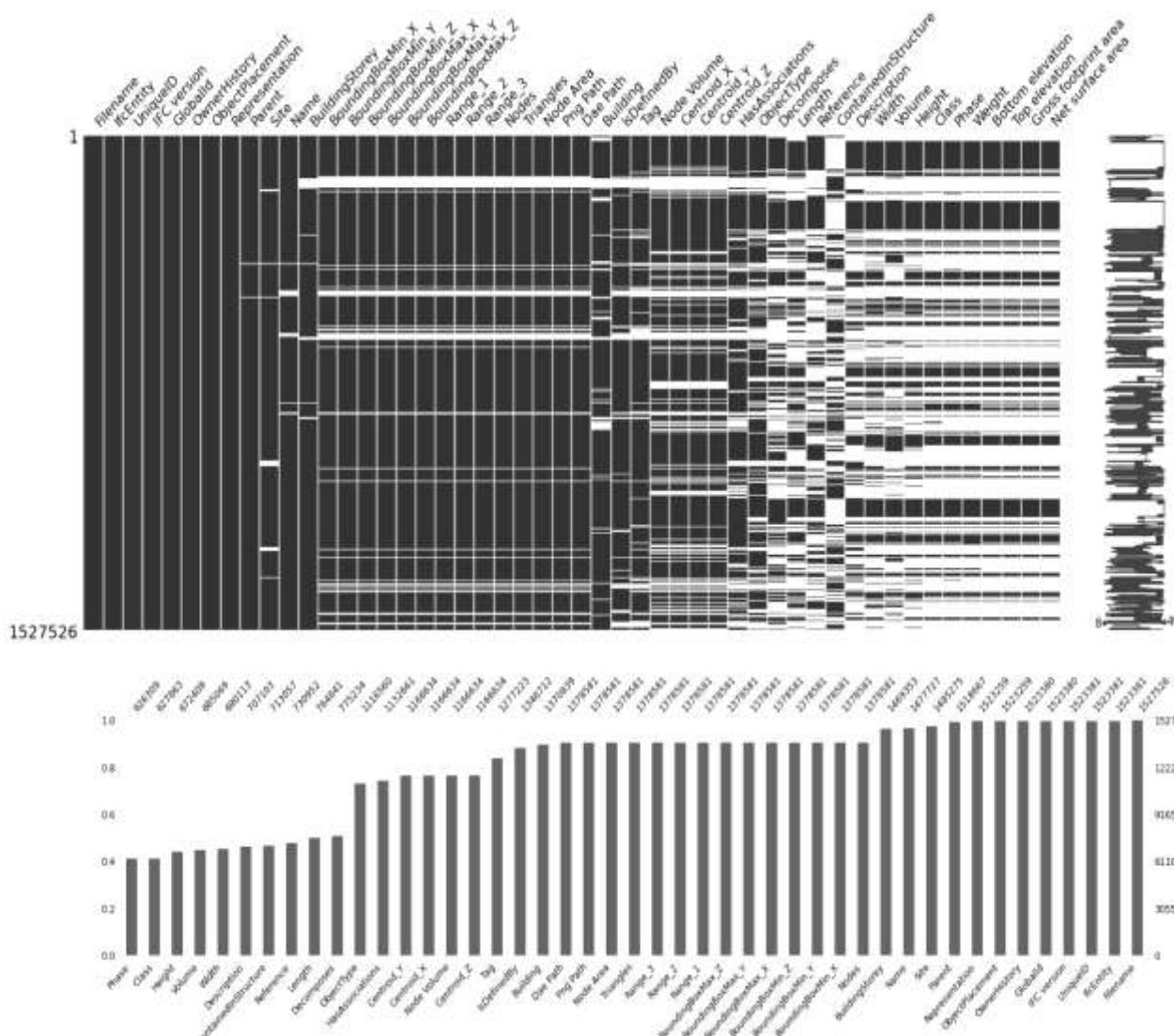
Po zhromaždení projektov vo formátoch RVT a IFC rôznych verzií a ich konverzii do štruktúrovaného formátu CSV pomocou bezplatného súboru SDK pre reverzné inžinierstvo sa takmer 10-tisíc projektov RVT a IFC zhromaždilo do jedného veľkého súboru tabuľky Apache Parquet a nahral sa do súboru Pandas DataFrame na analýzu (obrázok 9.1-9).



Obrázok 9.1-9 Štruktúrované údaje o projekte umožňujú spojiť ľubovoľný počet projektov do jednej dvojrozmernej tabuľky.

Údaje z tohto rozsiahleho súboru obsahujú tieto informácie: súbor súborov IFC obsahuje približne 4 milióny entít (riadkov) a 24 962 atribútov (stĺpcov) a súbor súborov RVT, ktorý pozostáva z približne 6 miliónov entít (riadkov), obsahuje 27 025 rôznych atribútov (stĺpcov).

Tieto súbory informácií (obrázok 9.1-10) zahŕňajú milióny prvkov, pre každý z nich boli dodatočne získané súradnice geometrie Bounding Box (obdĺžnik definujúci hranice objektu v projekte) a pridané do spoločnej tabuľky - súradnice geometrie Bounding Box (obdĺžnik definujúci hranice objektu v projekte) a vytvorené obrázky každého prvku vo formáte PNG a geometrie v otvorenom formáte XML - DAE (Collada).



Obr. 9.1-10 Sabinet 1,5 milióna prvkov a vizualizácia (knížnica missingno) obsadenosti prvých 100 atribútov vo forme histogramu.

Takto sme získali všetky informácie o desiatkach miliónov prvkov zo 4 596 projektov IFC a 6 471 projektov RVT, kde boli všetky atribúty - vlastnosti všetkých prvkov entít a ich geometria (Bounding Box) prevedené do štruktúrovanej podoby jednej tabuľky (DataFrame) (obr. 9.1-10 - údaje o populáciách dataframe sa zobrazujú ako histogramy).

Histogramy (obr. 9.1-10, obr. 9.2-6, obr. 9.2-7) vykreslené počas procesu analýzy umožňujú rýchle posúdenie hustoty údajov a frekvencie výskytu hodnôt v stípcach. To poskytuje prvý pohľad na rozloženie prvkov, prítomnosť odľahlých hodnôt a potenciálnu užitočnosť jednotlivých atribútov pri analýze a tvorbe modelov strojového učenia.

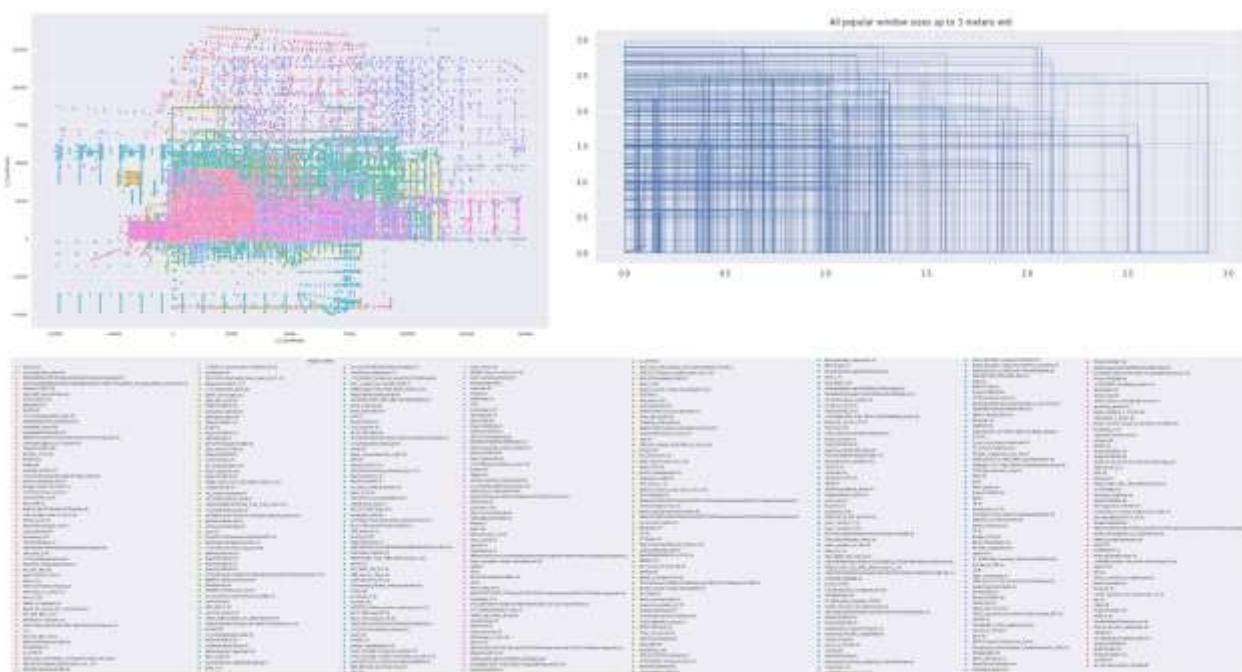
Jedným z príkladov praktického využitia tohto súboru údajov (obr. 9.1-10) je projekt "5000 IFC a RVT ". [149], ktorý je k dispozícii na platforme Kaggle. Predstavuje Jupyter Notebook s kompletným

riešením Pipeline: od predspracovania a analýzy údajov až po vizualizáciu výsledkov pomocou knižníc Pythonu - pandas, matplotlib, seaborn, folium a ďalších (obr. 9.1-11).



Obrázok 9.1-11 Príklady analýzy údajov z formátov CAD (BIM) pomocou vizualizačných knižníc Python a knižnice pandas.

Na základe metainformácií je možné určiť, v ktorých mestách boli vypracované určité projekty, a zobraziť ich na mape (napr. pomocou knižnice folium). Okrem toho časové pečiatky v údajoch umožňujú skúmať vzorce toho, kedy boli súbory uložené alebo upravené: podľa dňa v týždni, dennej doby a mesiaca.



Obr. 9.1-12 Vizualizácia geometrickej polohy všetkých stípov a rozmerov všetkých okien do 3 metrov v projektoch zo zoznamu v spodnej časti grafu.

Geometrické parametre vo forme Bounding Box extrahované z modelov sa tiež hodia na súhrnné analýzy. Napríklad na obrázku 9.1-12 sú zobrazené dva grafy: ľavý zobrazuje rozdelenie vzdialenosí medzi stĺpcami pre všetky projekty vzhľadom na nulový bod a pravý zobrazuje rozmery všetkých okien do výšky 3 m vo vzorke desiatok tisíc okenných prvkov (po zoskupení celého súboru údajov podľa parametra "Category" s hodnotou "OST\_Windows", "IfcWindows").

Kód analýzy Pipeline pre tento príklad a samotný súbor údajov sú k dispozícii na webovej stránke Kaggle pod názvom "5000 IFC a RVT | DataDrivenConstruction.io projects" [149]. Túto hotovú Pipeline spolu so súborom údajov možno skopírovať a spustiť zadarmo online na stránke Kaggle alebo offline v jednom z populárnych IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse s pluginom PyDev, Thonny, Wing IDE, IntelliJ IDEA s pluginom Python, JupyterLab alebo populárne online nástroje Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Analytické poznatky získané zo spracovania a štúdia obrovského množstva štruktúrovaných údajov budú zohrávať kľúčovú úlohu v rozhodovacích procesoch v stavebnictve.

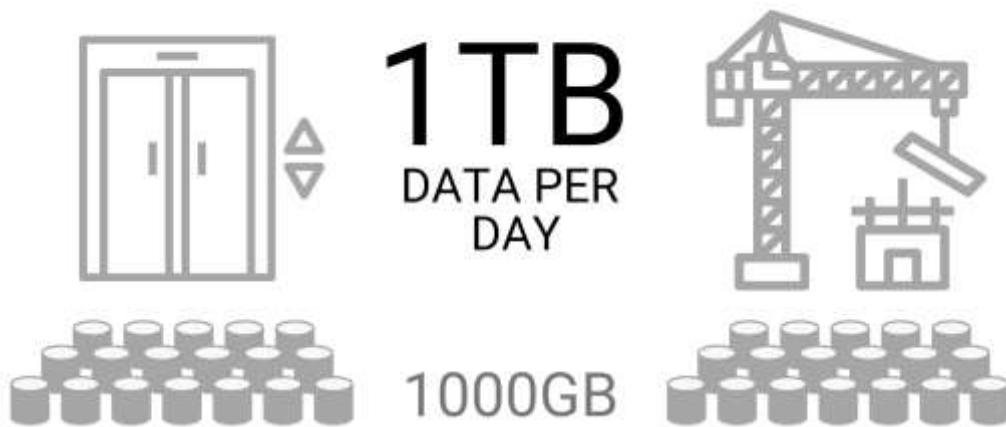
Vďaka takejto analýze informácií na základe predchádzajúcich projektov môžu odborníci efektívne predpovedať napríklad požiadavky na materiál a pracovnú silu a optimalizovať konštrukčné riešenia ešte pred začatím výstavby

Zatiaľ čo však údaje o projektoch alebo stavebných povoleniach sú relatívne statické informácie, ktoré sa menia pomerne pomaly, samotný proces výstavby sa rýchlo nasycuje rôznymi senzormi a zariadeniami internetu vecí: kamerami, automatizovanými monitorovacími systémami, ktoré prenášajú údaje v reálnom čase - to všetko mení stavenisko na dynamické digitálne prostredie, kde je potrebné analyzovať údaje v reálnom čase.

## **Internet vecí Internet vecí a inteligentné zmluvy**

Internet vecí Internet vecí predstavuje novú vlnu digitálnej transformácie, v ktorej každé zariadenie získava vlastnú IP adresu a stáva sa súčasťou globálnej siete. Internet vecí je koncept, ktorý zahrňa pripojenie fyzických objektov k internetu s cieľom zhromažďovať, spracovávať a prenášať údaje. V stavebnictve to znamená možnosť riadiť stavebné procesy v reálnom čase, minimalizovať plynvanie materiálom, predvídať opotrebovanie zariadení a automatizovať rozhodovanie.

Podľa článku CFMA "Príprava na budúcnosť s prepojenou výstavbou" [150] prejde stavebnictvo v nasledujúcom desaťročí veľkou digitálnou transformáciou, ktorá vyvrcholí konceptom Connected Construction - plne integrovaného a automatizovaného staveniska.



Obr. 9.1-13 IoT alebo dátové zariadenia na stavenisku môžu produkovať a prenášať terabajty dát denne.

Digitálne stavenisko znamená, že všetky prvky výstavby - od plánovania a logistiky až po realizáciu prác a kontrolu kvality na stavenisku pomocou pevných kamier a kvadrokoptér - budú integrované do jedného dynamického digitálneho ekosystému. Už skôr, v 7. časti tejto knihy, sme sa zaoberali možnosťami Apache NiFi (obrázok 7.4-5), bezplatného nástroja s otvoreným zdrojovým kódom, ktorý umožňuje streamovanie údajov v reálnom čase - od zberu z rôznych zdrojov až po prenos do úložísk alebo analytických platform.

Údaje o priebehu výstavby, spotrebe materiálu, stave zariadení a bezpečnosti sa budú v reálnom čase prenášať do analytických systémov (obr. 9.1-13). To umožňuje predvídať potenciálne riziká, promptne reagovať na odchýlky a optimalizovať procesy na stavbe. Medzi kľúčové komponenty digitálneho staveniska patria:

- IoT -senzory - sledovanie parametrov prostredia, monitorovanie stavebných zariadení a kontrola pracovných podmienok.
- Digitálne dvojčatá - virtuálne modely budov a infraštruktúry na predvídanie možných odchýlok a predchádzanie chybám.
- Automatizované logistické systémy - riadenie dodávateľského reťazca v reálnom čase s cieľom znížiť prestoje a náklady.
- Robotické stavebné systémy - používanie autonómnych strojov na vykonávanie rutinných a nebezpečných úloh.

Robotizácia používanie internetu vecí a koncepcia digitálneho staveniska (Connected Site) nielenže zvýšia efektivitu a znížia náklady, ale zároveň nastoľujú novú éru bezpečnosti, udržateľnej výstavby a prediktívneho riadenia projektov.

Tagy RFID (Radio Frequency Identification) sú tiež jednou z najdôležitejších súčasťí internetu vecí. Používajú sa na identifikáciu a sledovanie materiálov, strojov a dokonca aj pracovníkov na stavenisku, čím sa zvyšuje transparentnosť a kontrola zdrojov projektu.

RFID -technológia sa používa na automatické rozpoznávanie objektov pomocou rádiových signálov.

Pozostáva z troch kľúčových prvkov:

- RFID - štítky (pasívne alebo aktívne) - obsahujú jedinečný identifikátor a sú pripojené na materiály, nástroje alebo stroje.
- Skenery sú zariadenia, ktoré čítajú informácie zo štítkov a prenášajú ich do systému.
- Centralizovaná databáza - uchováva informácie o polohe, stave a pohybe objektov.

Použitie RFID v stavebníctve:

- Automatická evidencia materiálu - štítky na hotových betónových výrobkoch, výstuži alebo baleniach sendvičových panelov umožňujú kontrolu zásob a zabráňajú krádežiam.
- Kontrola práce personálu - RFID - zamestnanecké odznaky zaznamenávajú čas začiatku a konca zmeny, čím poskytujú záznam pracovného času.
- Monitorovanie zariadení - RFID - systém sleduje pohyb zariadení, čím predchádza prestojom a zvyšuje efektívnosť logistiky.

Tento balík technológií dopĺňajú inteligentné zmluvy založené na blockchaine, ktoré automatizujú platby, kontrolu dodávok a dodržiavanie zmlúv bez potreby sprostredkovateľov, čím sa znížuje riziko podvodov a oneskorení.

Kedže dnes neexistuje spoločný dátový model, inteligentné zmluvy sú jednoducho kódom, na ktorom sa účastníci dohodnú. Pomocou dátovo orientovaného prístupu je však možné vytvoriť spoločný model parametrov zmluvy, zakódovať ho do blockchainu a automatizovať plnenie podmienok.

Napríklad v systéme riadenia dodávateľského reťazca môže inteligentná zmluva sledovať doručenie zásielky na základe senzorov internetu vecí a štítkov RFID a automaticky previesť platbu, keď zásielka dorazí. Podobne na stavenisku môže inteligentná zmluva zaznamenať dokončenie pracovnej fázy - napríklad inštaláciu výstuže alebo vyliatie základov - na základe údajov z dronov alebo stavebných senzorov a automaticky iniciovať ďalšiu platbu dodávateľovi bez potreby manuálnych kontrol a papierových potvrdení.

Napriek novým technológiám a úsiliu medzinárodných normalizačných organizácií však množstvo konkurenčných nariem komplikuje situáciu v oblasti internetu vecí.

Podľa štúdie spoločnosti Cisco uverejnej v roku 2017 [151] sa takmer 60% iniciatív v oblasti internetu vecí (IoT) zastaví vo fáze overovania koncepcie a len 26% spoločností považuje svoje projekty internetu vecí za úplne úspešné. Okrem toho tretina dokončených projektov nedosahuje stanovené ciele a ani po realizácii nie je uznaná za úspešnú.

Jedným z hlavných dôvodov je nedostatočná interoperabilita medzi platformami, ktoré spracúvajú údaje z rôznych senzorov. V dôsledku toho zostávajú údaje izolované v rámci jednotlivých riešení. Alternatívou k tomuto prístupu, ako aj v iných podobných prípadoch (ktorým sme sa venovali v tejto knihe), je architektúra postavená na samotných údajoch ako primárnom aktíve.

Senzory internetu vecí zohrávajú kľúčovú úlohu nielen pri monitorovaní technického stavu zariadení, ale aj pri prediktívnej analýze, ktorá znižuje riziká na stavenisku a zlepšuje celkovú výkonnosť procesov predpovedaním porúch a odchýlok.

Údaje zozbierané snímačmi IoT a tagmi RFID môžu byť spracované v reálnom čase algoritmami strojového učenia, ktoré dokážu odhaliť anomálie a vopred upozorniť inžinierov na potenciálne poruchy. Môže ísť o rôzne situácie, od mikrotrhlín v betónových konštrukciách až po nezvyčajné prestávky v prevádzke vežového žeriavu, ktoré naznačujú technické poruchy alebo porušenie predpisov. Okrem toho môžu pokročilé algoritmy analýzy správania zachytiť vzorce správania, ktoré môžu naznačovať napríklad fyzickú únavu personálu, čím sa zlepší proaktívne riadenie bezpečnosti a pohody zamestnancov na stavenisku.

V stavebnictve sa nehody a poruchy - či už strojov alebo ľudí - zriedkavo stávajú náhle. Zvyčajne im predchádzajú drobné odchýlky, ktoré zostávajú nepovšimnuté. Prediktívna analýza a strojové učenie umožňujú odhaliť tieto signály v počiatočnom štádiu, ešte predtým, ako nastanú kritické následky.

Dokumenty, projektové súbory a údaje zo zariadení internetu vecí a štítkov RFID tvoria digitálnu stopu stavebných projektov, strojové učenie však môže pomôcť získať z nej užitočné poznatky. S nárastom objemu údajov a demokratizáciou prístupu k nim získava stavebnictvo nové možnosti v oblasti analytiky, prediktívnej analýzy a aplikácií umelej inteligencie.

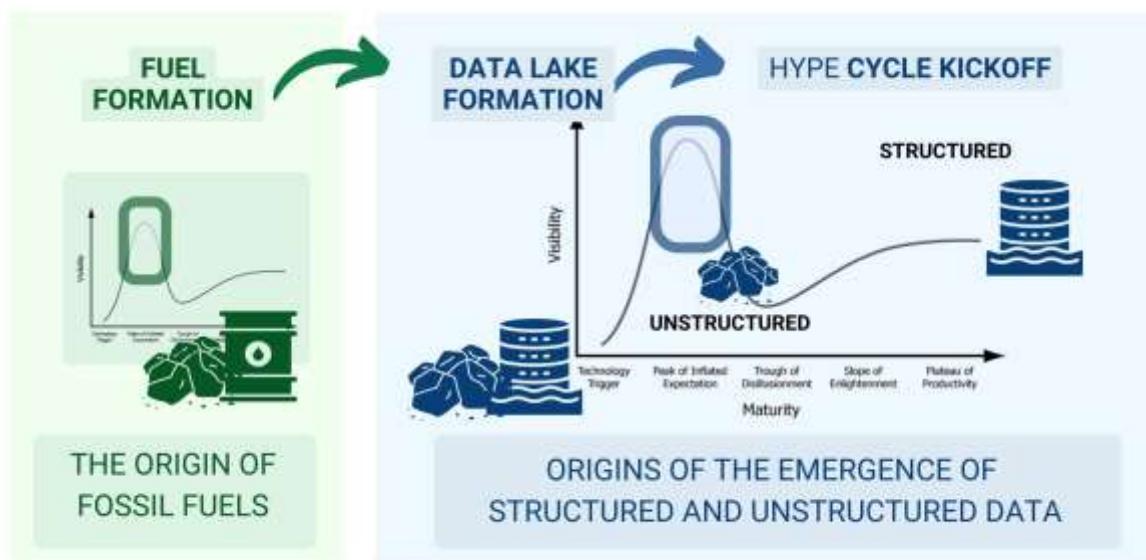


## KAPITOLA 9.2.

# STROJOVÉ UČENIE A PREDPOVEDE

### Strojové učenie a umelá inteligencia zmenia spôsob, akým tvoríme

Databázy rôznych systémov v stavebníctve - s ich nevyhnutne upadajúcou a čoraz zložitejšou infraštruktúrou - sa stávajú živnou pôdou pre budúce riešenia. Firemné servery sú podobne ako les bohaté na biomasu dôležitých informácií, často ukrytých pod zemou, v útrobách priečinkov a serverov. Masy údajov z rôznych systémov, ktoré sa dnes vytvárajú - po použití, po páde na dno servera a po rokoch fosilizácie - budú v budúcnosti poháňať strojové učenie a jazykové modely. Interné podnikové chaty (napr. samostatná inštancia lokálne nakonfigurovaného ChatGPT, LlaMa, Mistral, DeepSeek) budú postavené na týchto vnútropodnikových modeloch využívajúcich centralizované úložisko na rýchle a pohodlné získavanie informácií a vytváranie potrebných grafov, prehľadov a dokumentov.



Obr. 9.2-1 Tak ako sa stromy menia na uhlie, tak sa aj informácie pod tlakom času a analytyky časom menia na cennú obchodnú energiu.

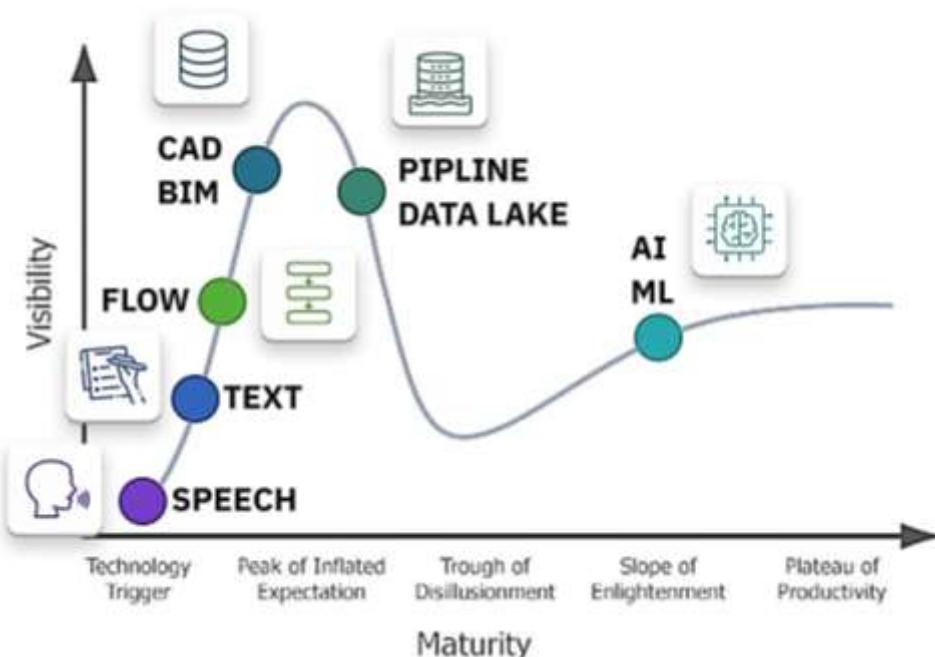
Fosilizácia rastlinnej hmoty v kombinácii s tlakom a teplotou vytvára homogénnu a jedinečne štruktúrovanú homogénnu hmotu stromov rôznych druhov, ktoré žili v rôznych časoch - drevené uhlie [152]. Rovnako aj informácie zaznamenané na pevných diskoch v rôznych formátoch a v rôznych časoch pod tlakom analytických oddelení a teplotou riadenia kvality nakoniec vytvoria homogénnu štruktúrovanú masu cenných informácií (obr. 9.2-1).

Tieto vrstvy (alebo častejšie izolované nugety) informácií vznikajú vďaka starostlivej organizácii údajov skúsenými analytikmi, ktorí zo zdanlivu dlho nepodstatných údajov začnú postupne získavať cenné informácie.

V okamihu, keď tieto vyspelé dátové vrstvy už nie sú len "vypálené" v reportoch, ale začnú obiehať v podnikových procesoch, obohacovať rozhodnutia a zlepšovať procesy, je podnik pripravený na ďalší

krok - prechod na strojové učenie a umelú inteligenciu (obr. 9.2-2).

Strojové učenie (ML - Machine learning) je trieda metód na riešenie problémov umelej inteligencie. Algoritmy strojového učenia rozpoznávajú vzory vo veľkých súboroch údajov a používajú ich na učenie sa. Každý nový súbor údajov umožňuje matematickým algoritmom zlepšovať sa a prispôsobovať podľa získaných informácií, čo umožňuje neustále zlepšovať presnosť odporúčaní a predpovedí.



Obr. 9.2-2 Zanikajúce technológie tvorby dát a aplikácia analytických nástrojov otvárajú dvere téme strojového učenia.

Ako povedal vplyvný generálny riaditeľ najväčšieho investičného fondu na svete (ktorý vlastní kľúčové podiely v takmer všetkých najväčších spoločnostiach vyrábajúcich stavebný softvér, ako aj v spoločnostiach, ktoré vlastnia najväčšie množstvo nehnuteľností na svete [55]) v rozhovore z roku 2023 - strojové učenie zmení svet stavebníctva.

*AI má obrovský potenciál. Zmení spôsob, akým pracujeme a žijeme. Umelá inteligencia a robotika zmenia spôsob, akým pracujeme a ako stavíame, a budeme môcť využiť umelú inteligenciu a robotiku ako prostriedok na vytvorenie oveľa vyššej produktivity [153].*

- Generálny riaditeľ najväčšieho investičného fondu na svete, rozhovor, september 2023.

Strojové učenie (ML) funguje na základe spracovania veľkého množstva údajov pomocou štatistických techník, ktoré napodobňujú aspekty ľudského myslenia. Väčšina spoločností však nemá takéto súbory

údajov, a ak ich má, často nie sú dostatočne označené. Tu môžu pomôcť sémantické technológie a transferové učenie, technika, ktorá umožňuje ML efektívnejšie pracovať s malým množstvom údajov, o ktorej realizovateľnosti sme hovorili v predchádzajúcich kapitolách tejto časti.

Podstata transferového učenia spočíva v tom, že namiesto toho, aby ste sa každú úlohu učili od začiatku, môžete využiť znalosti získané v príbuzných oblastiach. Je potrebné si uvedomiť, že vzory a objavy z iných odvetví sa dajú prispôsobiť a aplikovať v stavebníctve. Napríklad metódy optimalizácie logistických procesov vyvinuté v maloobchode pomáhajú zvýšiť efektívnosť riadenia dodávateľského reťazca v stavebníctve. Analýza veľkých dát, ktorá sa aktívne využíva vo finančníctve, sa môže aplikovať na prognózovanie nákladov a riadenie rizík v stavebných projektoch. A technológie počítačového videnia a robotiky, ktoré sa vyvíjajú v priemysle, už nachádzajú uplatnenie pri automatizovanej kontrole kvality, monitorovaní bezpečnosti a riadení zariadení staveniska.

Transferové učenie umožňuje nielen urýchliť zavádzanie inovácií, ale aj znížiť náklady na ich vývoj s využitím už získaných skúseností z iných odvetví.

$$\text{labor productivity in construction} = f(\text{AI})$$

Obr. 9.2-3 Umelá inteligencia technológie a robotika budú v budúcnosti hlnacou silou zvyšovania produktivity v stavebníctve.

Ľudské myšlenie je organizované na podobnom princípe: pri riešení nových problémov vychádzame z predtým získaných vedomostí (obr. 4.4-19, obr. 4.4-20, obr. 4.4-21). V strojovom učení tento prístup funguje tiež - zjednodušením dátového modelu a jeho elegantnejším spracovaním môžeme znížiť zložitosť problému pre algoritmy ML. To zase znižuje potrebu veľkého množstva údajov a znižuje výpočtové náklady.

### **Od subjektívneho hodnotenia k štatistickej prognóze**

Éra, keď strategické rozhodnutia záviseli na intuícii jednotlivých manažérov (obrázok 9.2-4), je už minulosťou. V čoraz konkurenčnejšom a náročnejšom ekonomickom prostredí sa subjektívny prístup stáva príliš riskantným a neefektívnym. Podniky, ktoré sa naďalej spoliehajú na osobné názory namiesto objektívnej analýzy údajov, strácajú schopnosť rýchlo reagovať na zmeny.

Konkurenčné prostredie si vyžaduje presnosť a opakovateľnosť na základe údajov, štatistických vzorcov a vypočítateľnej pravdepodobnosti. Rozhodnutia už nemôžu byť založené na pocitoch, musia byť založené na koreláciách, trendoch a prediktívnych modeloch odvodených z analytiky a strojového učenia. Nejde len o zmenu nástrojov - je to zmena logiky myslenia: od predpokladov k dôkazom, od subjektívnych pravdepodobností k štatisticky vypočítaným odchýlkam, od pocitov k faktom.



Obr. 9.2-4 Éra rozhodovania podľa HiPPO (názor najlepšie plateného zamestnanca) sa s nástupom veľkých dát a strojového učenia stane minulosťou.

Manažéri, ktorí sa spoliehali len na svoje vlastné pocity, budú nevyhnutne čeliť novej realite: autorita už nerozhoduje o výbere. V centre riadenia sú teraz systémy, ktoré analyzujú milióny parametrov a vektorov, identifikujú skryté vzorce a navrhujú optimálne stratégie.

Hlavným dôvodom, prečo sa spoločnosti dnes stále vyhýbajú implementácii ML, je jeho nedostatočná transparentnosť. Väčšina modelov funguje pre manažérov ako "čierne skrinky" bez vysvetlenia, ako presne dospejú k svojim záverom. To vedie k problémom: algoritmy môžu posilňovať stereotypy a dokonca vytvárať humorné situácie, ako v prípade chatbota spoločnosti Microsoft, ktorý sa rýchlo zmenil na toxickej komunikačnej nástroj [154].

Garry Kasparov, bývalý majster sveta v šachu, sa v knihe Deep Thinking zamýšľa nad svojou porážkou od počítača IBM Big Blue [155]. Tvrdí, že skutočná hodnota umelej inteligencie nie je v kopírovaní ľudskej inteligencie, ale v dopĺňaní našich schopností. AI by mala vykonávať úlohy, v ktorých sú ľudia slabí, zatiaľ čo ľudia prinášajú kreativitu. Počítače zmenili tradičný prístup k analýze šachu. Namesto vytvárania fascinujúcich príbehov o partiach počítačové šachové programy hodnotia každý ťah nestranne, len na základe jeho skutočnej sily alebo slabosti. Kasparov poznamenáva, že ľudská tendencia vnímať udalosti ako ucelené príbehy, a nie ako jednotlivé akcie, často vedie k nesprávnym záverom - nielen v šachu, ale v živote všeobecne.

Preto ak plánujete používať strojové učenie na predpovedanie a analýzu, je dôležité pochopiť jeho základné princípy - ako fungujú algoritmy a ako sa spracovávajú údaje - skôr, ako začnete pri svojej práci používať nástroje strojového učenia a AI. Najlepším spôsobom, ako začať, sú praktické skúsenosti.

Jedným z najvhodnejších nástrojov na úvodné zoznámenie sa s tému strojového učenia a predikcie je Jupyter Notebook a populárny klasický súbor údajov Titanic, ktorý poskytne vizuálny úvod do kľúčových metód analýzy údajov a tvorby modelov ML.

## Súbor údajov Titanic: Hello World vo svete analytiky data and big data

Jedným z najznámejších príkladov použitia ML v dátovej analytike je analýza súboru údajov z Titaniku, ktorý sa často používa na štúdium pravdepodobnosti prežitia cestujúcich. Učenie tejto tabuľky je analogické programu "Hello World" pri učení sa programovacích jazykov.

Pri potopení lode RMS Titanic v roku 1912 zahynulo 1502 z 2224 ľudí. Súbor údajov o Titaniku obsahuje nielen informácie o tom, či cestujúci prežil, ale aj atribúty ako: vek, pohlavie, trieda letenky a ďalšie parametre. Tento súbor údajov je k dispozícii bezplatne a možno ho otvoriť a analyzovať na rôznych offline a online platformách.

Odkaz na súbor údajov o Titaniku:

<https://raw.githubusercontent.com/datasets/master/titanic.csv>

Už skôr v kapitole "IDE s podporou LLM a budúce zmeny v programovaní" sme sa venovali Jupyter Notebook - jednému z najpopulárnejších vývojových prostredí na analýzu údajov a strojové učenie. Bezplatnou cloudrovou obdobou Jupyter Notebook sú platformy Kaggle a Google Collab, ktoré umožňujú spúštať kód v jazyku Python bez inštalácie softvéru a poskytujú bezplatný prístup k výpočtovým zdrojom.

Kaggle je najväčšia súťažná platforma pre analýzu údajov a strojové učenie s integrovaným prostredím na vykonávanie kódu. Od októbra 2023 má Kaggle viac ako 15 miliónov používateľov [156] zo 194 krajín.

Stiahnite si a použite súbor údajov Titanic na platforme Kaggle (obrázok 9.2-5) na uloženie súboru

údajov (jeho kópie) a spustenie kódu Python s predinštalovanými knižnicami priamo v prehliadači bez nutnosti inštalácie špecializovaného vývojového prostredia.



Obr. 9.2-5 Štatistiky tabuľky Titanic - najobľúbenejšia trénovacia množina údajov na výučbu dátovej analýzy a strojového učenia.

Súbor údajov o Titaniku obsahuje údaje o 2 224 cestujúcich na palube *lode RMS Titanic* v čase jej stroskotania v roku 1912. Súbor údajov je prezentovaný ako dve samostatné tabuľky, tréningová (train.csv) a testovacia (test.csv) vzorka, čo umožňuje jeho použitie na trénovanie modelov aj na vyhodnotenie ich presnosti na nových údajoch.

Tréningový súbor údajov obsahuje atribúty - atribúty cestujúcich (vek, pohlavie, trieda letenky a iné) a informácie o tom, kto prežil (stĺpec s binárnymi hodnotami "Prežil"). Trénovací súbor údajov (obrázok 9.2-6 - súbor train.csv) sa používa na trénovanie modelu. Testovací súbor údajov (obr. 9.2-7 - súbor test.csv) obsahuje len atribúty cestujúcich bez informácií o preživších (bez jediného stĺpca "Prežil"). Testovací súbor údajov je určený na testovanie modelu na nových údajoch a na vyhodnotenie jeho presnosti.

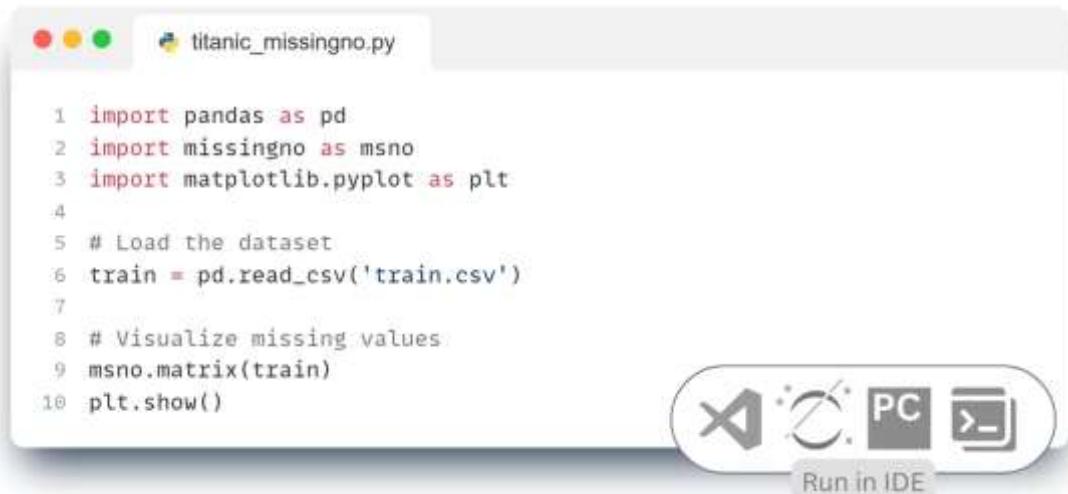
Takto máme takmer identické atribúty cestujúcich v tréningových a testovacích súboroch údajov. Jediný kľúčový rozdiel je v tom, že v testovacom súbore údajov máme zoznam cestujúcich, ktorí nemajú stĺpec "Survivor" - cieľovú premennú, ktorú sa chceme naučiť predpovedať pomocou rôznych matematických algoritmov. A po zostavení modelu budeme môcť porovnať výstup nášho modelu so skutočným parametrom "Survivor" z testovacieho súboru údajov, ktorý budeme brať do úvahy pri vyhodnocovaní výsledkov.

Hlavné stĺpce tabuľky, parametre cestujúcich v tréningovom a testovacom súbore údajov:

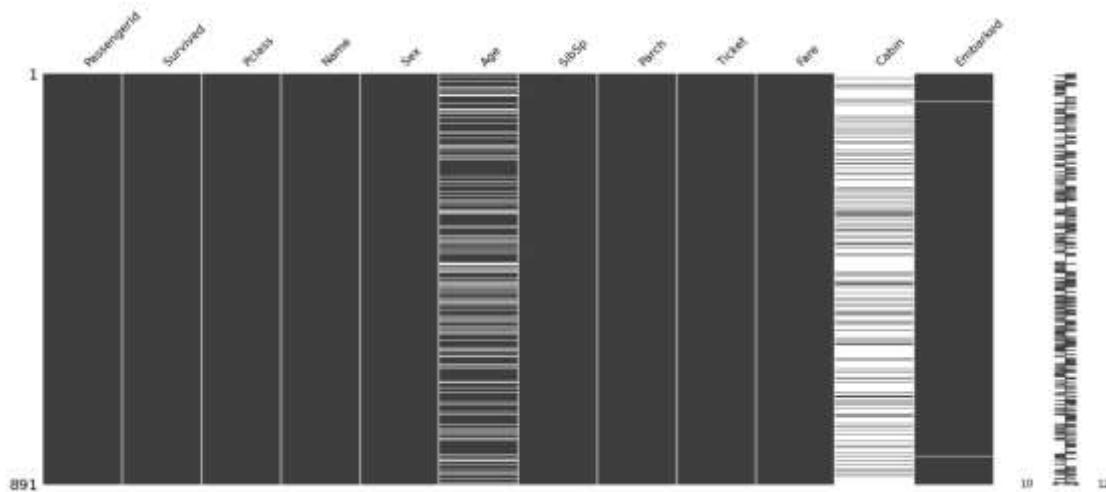
- **PassengerId** - jedinečný identifikátor cestujúceho
- **Prežil** - 1, ak cestujúci prežil, 0, ak je mŕtvy (nie je k dispozícii v testovacom súbore)
- **Pclass** - trieda lístka (1, 2 alebo 3)

- **Meno** - meno cestujúceho
- **Pohlavie** - pohlavie cestujúceho (muž/žena)
- Vek
- **SibSp** - počet súrodencov alebo manželov na palube
- **Parch** - počet rodičov alebo detí na palube
- **Vstupenka** - číslo vstupenky
- **Cestovné** - cena lístka
- Kabína - číslo kabíny (mnohé údaje chýbajú)
- **Nalodenie** je prístav naloženia (C = Cherbourg, Q = Queenstown, S = Southampton).

Na vizualizáciu chýbajúcich údajov v oboch tabuľkách môžete použiť knižnicu missingno (obr. 9.2-6, obr. 9.2-7), ktorá zobrazuje chýbajúce hodnoty vo forme histogramu, kde biele polia označujú chýbajúce údaje. Táto vizualizácia umožňuje rýchle posúdenie kvality údajov pred ich spracovaním.



```
1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()
```



Obr. 9.2-6 Niekoľko riadkov kódu sa používa na vizualizáciu chýbajúcich údajov v súbore trénovalých údajov Titanic, kde kľúčovým parametrom pre trénovanie je parameter "Survived" (prežil).



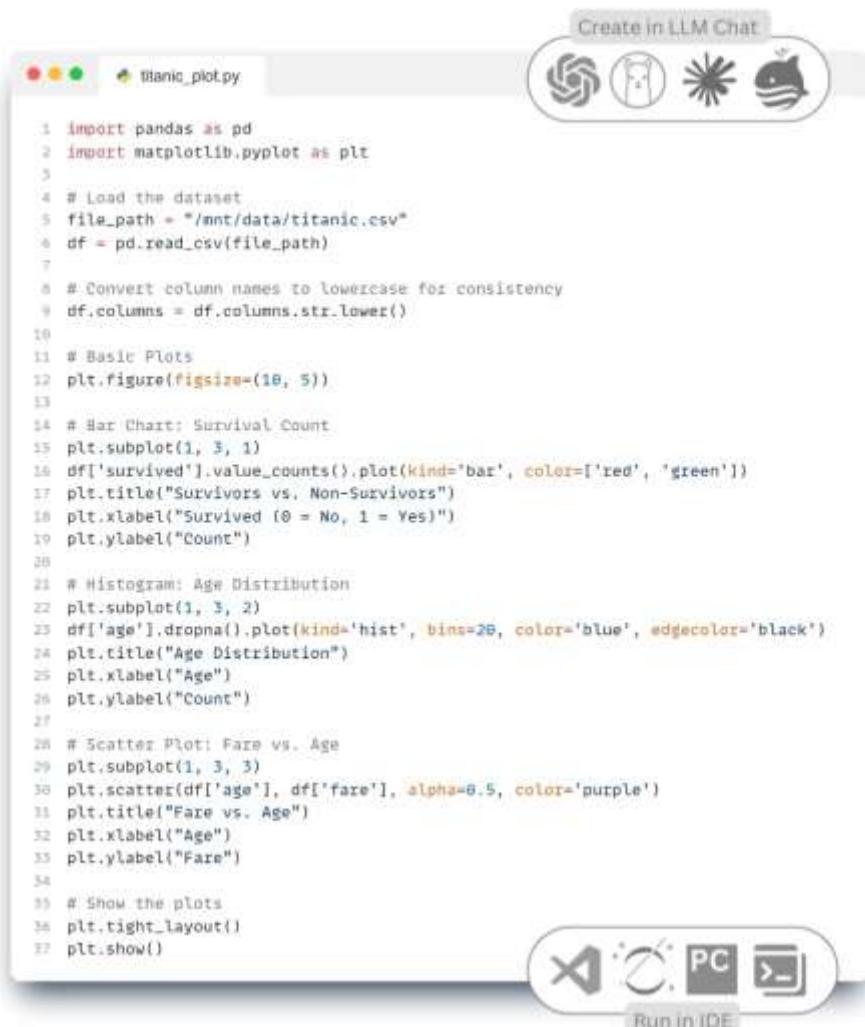
Obrázok 9.2-7 Vizualizácia chýbajúcich údajov v testovacom súbore údajov Titanic, ktorý obsahuje len charakteristiky cestujúcich bez informácií.

Pred formulovaním hypotéz a vytváraním predpovedí na základe súboru údajov pomáhajú vizuálne analýzy identifikovať kľúčové vzory v údajoch, posúdiť ich kvalitu a určiť možné závislosti. Existuje mnoho vizualizačných techník, ktoré vám môžu pomôcť lepšie pochopíť súbor údajov Titanic. Na posúdenie kvality informácií a pochopenie údajov môžete použiť distribučné grafy na analýzu vekových skupín cestujúcich, grafy prežitia podľa pohlavia a triedy a matice chýbajúcich údajov.

- ❷ Požiadajme LLM, aby nám pomohol vizualizovať údaje zo súboru údajov Titanic zaslaním nasledujúcej textovej požiadavky ľubovoľnému modelu LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN alebo akémukoľvek inému):

Ukážte prosím niekoľko jednoduchých grafov pre súbor údajov o Titaniku. Sami si stiahnite súbor údajov a ukážte ↗

- ❸ Odozva LLM vo forme pripraveného kódu a grafov vizualizujúcich parametre súboru údajov

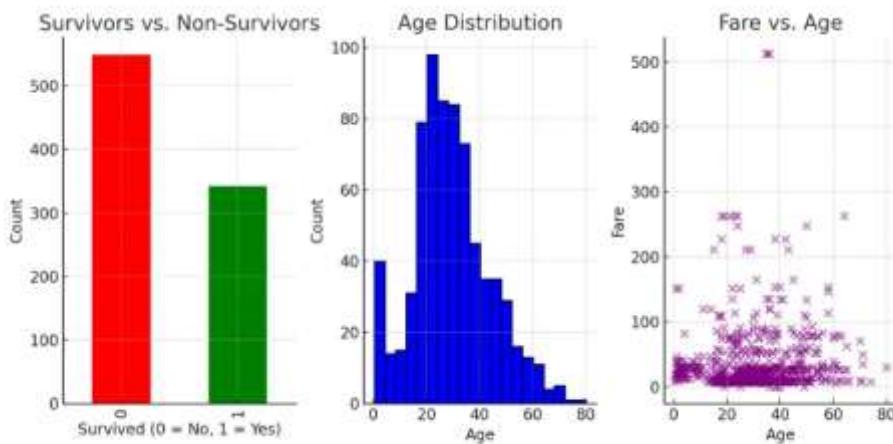


```

Create in LLM Chat
titanic_plot.py
  1 import pandas as pd
  2 import matplotlib.pyplot as plt
  3
  4 # Load the dataset
  5 file_path = "/mnt/data/titanic.csv"
  6 df = pd.read_csv(file_path)
  7
  8 # Convert column names to lowercase for consistency
  9 df.columns = df.columns.str.lower()
 10
 11 # Basic Plots
 12 plt.figure(figsize=(10, 5))
 13
 14 # Bar Chart: Survival Count
 15 plt.subplot(1, 3, 1)
 16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
 17 plt.title("Survivors vs. Non-Survivors")
 18 plt.xlabel("Survived (0 = No, 1 = Yes)")
 19 plt.ylabel("Count")
 20
 21 # Histogram: Age Distribution
 22 plt.subplot(1, 3, 2)
 23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
 24 plt.title("Age Distribution")
 25 plt.xlabel("Age")
 26 plt.ylabel("Count")
 27
 28 # Scatter Plot: Fare vs. Age
 29 plt.subplot(1, 3, 3)
 30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
 31 plt.title("Fare vs. Age")
 32 plt.xlabel("Age")
 33 plt.ylabel("Fare")
 34
 35 # Show the plots
 36 plt.tight_layout()
 37 plt.show()

```

Run in IDE



Obrázok 9.2-8 LLM vám pomôže získať okamžitú vizualizáciu údajov vášho súboru údajov.

Vizualizácia údajov je dôležitým krokom na prípravu súboru údajov na následnú konštrukciu modelu strojového učenia, ku ktorému sa dá dostať len na základe pochopenia údajov.

### Strojové učenie v praxi: od pasažierov Titaniku po riadenie projektov

Hlavnou hypotézou použitou na skúmanie rámca strojového učenia na základe súboru údajov z Titaniku je, že určité skupiny cestujúcich mali väčšiu šancu na prežitie.

Malá tabuľka pasažierov Titaniku sa stala populárnu na celom svete a milióny ľudí ju používajú na trénovanie, experimentovanie a testovanie modelov, aby zistili, ktoré algoritmy a hypotézy vytvoria najpresnejší model predpovede prežitia na základe tréningového súboru údajov pre pasažierov Titaniku.

Príťažливosť súboru údajov o Titaniku spočíva v jeho kompaktnosti: s niekoľkými stovkami riadkov a dvanásťimi stĺpcami (obr. 9.2-6) poskytuje dostatok príležitostí na analýzu. Súbor údajov je relatívne jednoducho klasickým príkladom riešenia binárnej klasifikácie, kde cieľ problému - prežitie - je vyjadrený v pohodlnom formáte 0 alebo 1.

John Wheeler vo filme "It from Bit" [7] tvrdí, že vesmír je založený na binárnych voľbách. Podobne aj podnik vedený ľuďmi zloženými z molekúl je v skutočnosti postavený na sérii binárnych binárnych volieb.

Okrem toho sú údaje založené na skutočnej historickej udalosti, čo ich na rozdiel od umelo vytvorených príkladov robí cennými pre výskum. Len na platforme Kaggle, ktorá je jednou z najväčších platform Data Pipeline a ETL, sa na výzvach založených na súbore údajov o Titaniku zúčastnilo 1 355 998 ľudí, ktorí vytvorili 53 963 jedinečných riešení Data Pipeline [157] (obrázok 9.2-9).

Zdá sa to neuveriteľné, ale len 1000 riadkov údajov o pasažieroch Titaniku s 12 parametrami sa stalo poľom pre milióny hypotéz, logických reťazcov a jedinečných dátových potrubí. Z malého súboru údajov sa rodia nekonečné poznatky, hypotézy a interpretácie - od jednoduchých modelov prežitia až po komplexné súbory, ktoré zohľadňujú skryté zákonitosti a zložité labyrinty uvažovania.

Rank	Title	Updated	Comments	Score	Status
1	Titanic Tutorial	3y ago	29858	16916	Gold
2	Titanic competition w/ TensorFlow Decision Forests	2y ago	318	1098	Gold
3	Titanic Data Science Solutions	6y ago	2590	10723	Gold
4	Exploring Survival on the Titanic	7y ago	1072	3968	Gold

Obr. 9.2-9 Prvých päť riešení z celkového počtu 53 963 hotových a otvorených riešení Pipeline. Len na portáli Kaggle [157] sa tento problém pokúsilo vyriešiť už takmer 1,5 milióna ľudí.

Ak aj takáto malá tabuľka dokáže vygenerovať milióny unikátnych riešení (obr. 9.2-9), čo môžeme povedať o skutočných súboroch priemyselných konštrukčných údajov, kde sa parametre merajú v desiatkach tisíc?

Štandardný CAD -návrh relatívne malej budovy obsahuje desiatky tisíc entít s tisícami parametrov - od geometrických charakteristik až po atribúty nákladov a času. Predstavte si, koľko potenciálnych poznatkov, vzťahov, predpovedí a hypotéz o riadení sa skrýva v údajoch zo všetkých projektov vašej spoločnosti zozbieraných v priebehu rokov. Historické údaje o projektoch nie sú len archívom - je to živá pamäť organizácie, jej digitálna stopa, ktorú možno analyzovať a vytvoriť tak veľké množstvo jedinečných hypotéz.

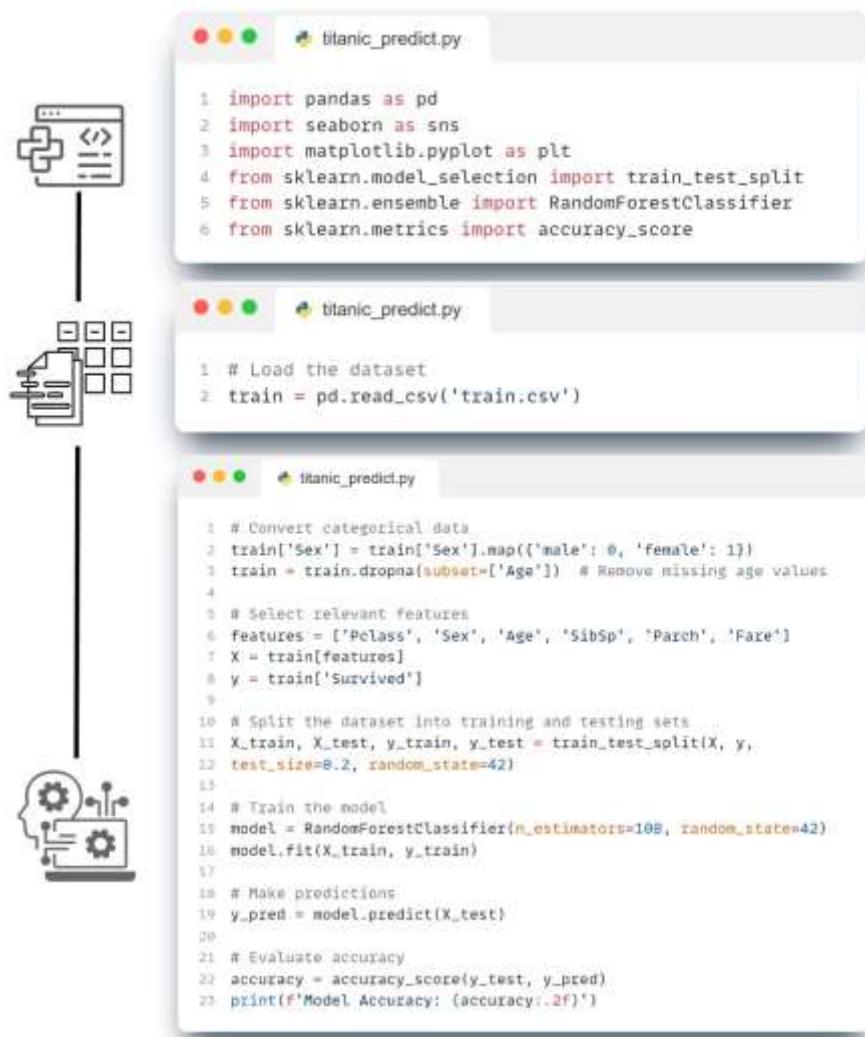
Najdôležitejšie je, že nemusíte čakať, kým sa komunita Kaggle začne zaujímať o vašu spoločnosť alebo vaše údaje. Môžete začať pracovať s tým, čo máte k dispozícii už dnes: spúštať analýzy na vlastných

údajoch, trénovať modely na vlastných údajoch, identifikovať opakovania, anomálie a vzory. Tam, kde sa kedysi vyžadovali roky experimentovania a drahé konzultácie, dnes stačí iniciatíva, titul LLM, otvorený prístup k údajom a ochota učiť sa.

- ❷ Na vytvorenie algoritmu strojového učenia, ktorý bude predpovedať mieru prežitia cestujúcich na základe súboru trénovacích údajov o cestujúcich train.csv, požiadajme LLM, aby tento problém vyriešil za nás:

Na základe tréningového súboru údajov o pasažieroch Titaniku vytvorte model strojového učenia na predpovedanie schopnosti prežiť. ↵

💡 Odpoveď LLM :



```

titanic_predict.py
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

# Load the dataset
train = pd.read_csv('train.csv')

# Convert categorical data
train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
train = train.dropna(subset=['Age']) # Remove missing age values

# Select relevant features
features = ['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare']
X = train[features]
y = train['Survived']

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Train the model
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Model Accuracy: {accuracy:.2f}')

```

Obr. 9.2-10 LLM zostavil predpoveď prežívších Titaniku pomocou algoritmu strojového učenia Random Forest.

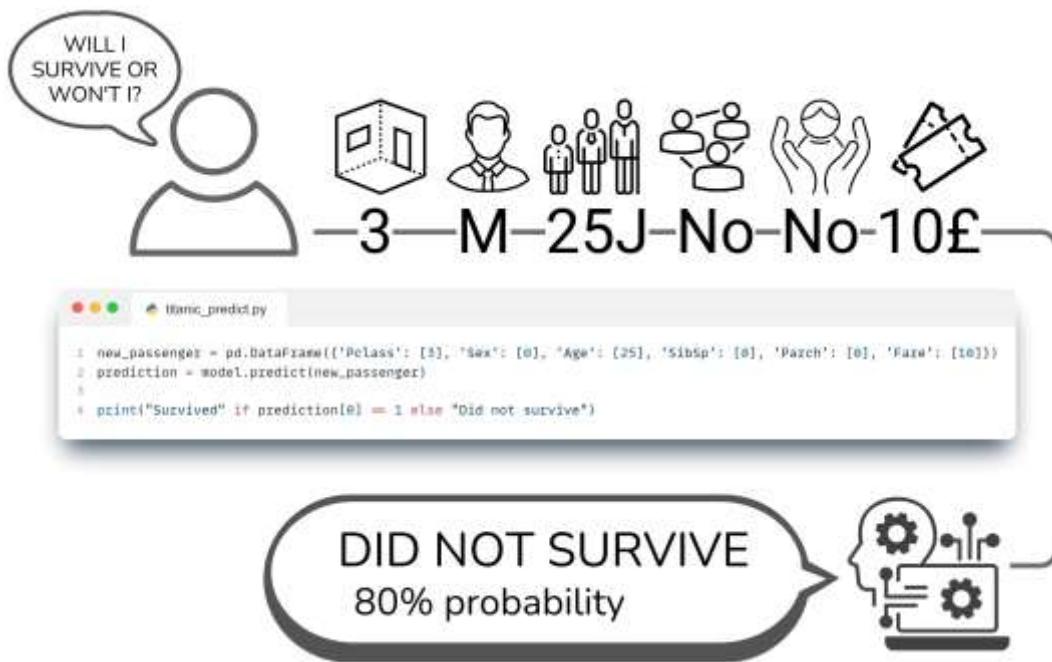
Výsledný kód z LLM (obr. 9.2-10) načíta údaje o pasažieroch Titaniku, vyčistí ich, konvertuje kategoriálne premenné (napr. pohlavie do číselného formátu) a trénuje model prostredníctvom algoritmu RandomForestClassifier na predpovedanie, či pasažier prežil alebo nie (o populárnych algoritmoch budeme hovoriť viac v nasledujúcich kapitolách).

Kód rozdelí trénované údaje na trénovaciu a testovaciu množinu (na webovej stránke Kaggle sú už vytvorené hotové súbory test.csv (obr. 9.2-7) a train.csv (obr. 9.2-6) na trénovanie, potom sa model trénuje na trénovaných údajoch a testuje na testovacích údajoch, aby sa zistilo, aký dobrý je konkrétny predikčný model. Po vyškolení sa do modelu vložia testovacie údaje z test.csv (so skutočnými údajmi

o tých, ktorí prežili alebo neprežili) a model predpovedá, kto prežil a kto nie. V našom prípade je presnosť modelu strojového učenia, ktorý sme získali, približne 80%, čo svedčí o tom, že pomerne dobre zachytáva vzory.

Strojové učenie možno prirovnáť k dieťaťu, ktoré sa snaží vložiť obdĺžnikový kváder do okrúhleho otvoru. V počiatočných fázach algoritmus skúša mnoho prístupov, pričom naráža na chyby a nezrovnalosti. Tento proces sa môže zdať neefektívny, ale poskytuje dôležité učenie: analýzou každej chyby model zlepšuje svoje predpovede a robí čoraz presnejšie rozhodnutia.

Teraz možno tento model (obr. 9.2-10) použiť na predpovedanie miery prežitia nových cestujúcich a napríklad, ak mu dodáte informácie o cestujúcich pomocou funkcie `model.predict` s parametrami: "muž", "3. trieda", "25 rokov", "bez príbužných na palube", model vytvorí predpoved - že cestujúci s 80% pravdepodobnosťou neprežije katastrofu, ak bol na Titaniku v roku 1912 (obr. 9.2-11).



Obrázok 9.2-11 Model, ktorý sme vytvorili vyššie, teraz dokáže s 80% pravdepodobnosťou predpovedať, či niektorý nový pasažier Titaniku prežije.

Model predpovede prežitia pasažierov Titaniku ilustruje oveľa širší koncept: tisíce odborníkov v stavebnictve robia každý deň podobné "dvojité" rozhodnutia - o živote alebo smrti rozhodnutia, projektu, odhadu, nástroja, zisku alebo strate, bezpečnosti alebo riziku. Podobne ako v príklade Titaniku, kde výsledok závisel od faktorov (pohlavie, vek, trieda), v stavebnictve každý aspekt rozhodnutia ovplyvňuje množstvo vlastných faktorov a premenných (stĺpce tabuľiek): náklady na materiál, kvalifikácia pracovníkov, termíny, počasie, logistika, technické riziká, pripomienky a stáť sice ďalších parametrov.

V stavebníctve sa strojové učenie riadi rovnakými princípmi ako v iných oblastiach: modely sa trénujú na historických údajoch - z projektov, zákaziek, odhadov - s cieľom testovať rôzne hypotézy a nájsť najefektívnejšie riešenia. Tento proces sa podobá výučbe dieťaťa metódou pokusov a omylov: s každým cyklom sa modely prispôsobujú a stávajú sa presnejšími.

Využívanie nahromadených údajov otvára nové obzory v oblasti stavebníctva. Namiesto prácnych manuálnych výpočtov možno trénovať modely, ktoré dokážu s vysokou presnosťou predpovedať kľúčové charakteristiky budúcich projektov. Prediktívna analyтика tak mení stavebníctvo na priestor, v ktorom možno nielen plánovať, ale aj s istotou predvídať vývoj.

## Predpoveď a prognózy na základe historických údajov

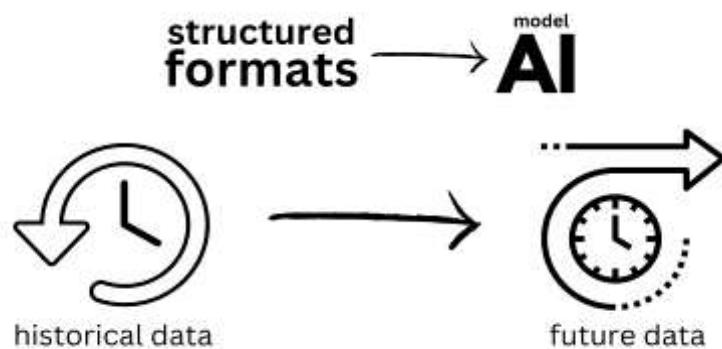
Údaje zozbierané o projektoch spoločnosti otvárajú možnosť vytvárať modely schopné predpovedať nákladové a časové charakteristiky budúcich, ešte nerealizovaných objektov - bez časovo náročných manuálnych výpočtov a porovnaní. To umožňuje výrazne urýchliť a zjednodušiť procesy oceňovania, pričom sa nespolieha na subjektívne predpoklady, ale na spoľahlivé matematické prognózy.

Predtým, vo štvrtej časti knihy, sme sa podrobne zaoberali tradičnými metódami odhadu nákladov na projekt vrátane metódy založenej na zdrojoch, a tiež sme spomenuli parametrické a expertné prístupy. Tieto metódy sú stále relevantné, ale v modernej praxi sa začínajú obohacovať o nástroje štatistickej analýzy a strojového učenia, ktoré môžu výrazne zlepšiť presnosť a reprodukovateľnosť odhadov.

Procesy manuálneho a poloautomatického výpočtu cien a časových atribútov budú v budúcnosti doplnené o názory a predpovede modelov ML, ktoré dokážu analyzovať historické údaje, nájsť skryté vzorce a navrhnuť informované rozhodnutia.

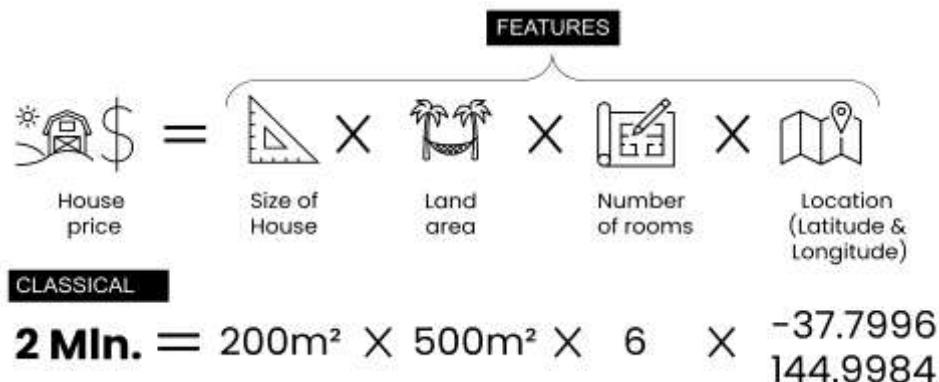
Nové údaje a scenáre sa budú generovať automaticky z už dostupných informácií - podobne ako jazykové modely (LLM) vytvárajú texty, obrázky a kódy na základe údajov zozbieraných v priebehu rokov z otvorených zdrojov [158].

Tak ako sa dnes ľudia pri posudzovaní budúcich udalostí spoliehajú na skúsenosti, intuíciu a interné štatistiky, v nasledujúcich rokoch bude budúcnosť stavebných projektov čoraz viac určovať kombinácia nahromadených znalostí a matematických modelov strojového učenia.



Obr. 9.2-12 Kvalitatívne a štruktúrované historické údaje o spoločnosti sú materiálom, na ktorom sú postavené modely strojového učenia a predpovede.

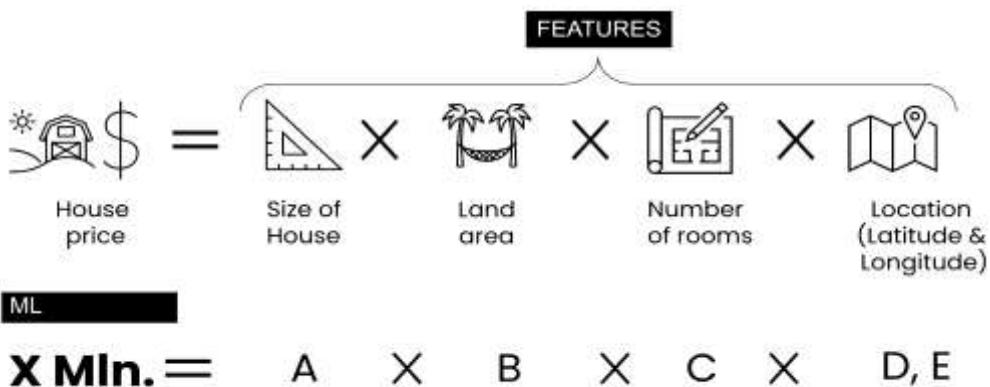
Zoberme si jednoduchý príklad: predpovedanie ceny domu na základe jeho rozlohy, veľkosti pozemku, počtu izieb a zemepisnej polohy. Jedným z prístupov je vytvorenie klasického modelu, ktorý analyzuje tieto parametre a vypočíta očakávanú cenu (obrázok 9.2-13). Tento prístup si vyžaduje presný a vopred známy vzorec, čo je v reálnej praxi prakticky nemožné.



Obr. 9.2-13 Klasický algoritmus možno použiť na odhad hodnoty domu s pevným vzorcом, ktorý treba nájsť.

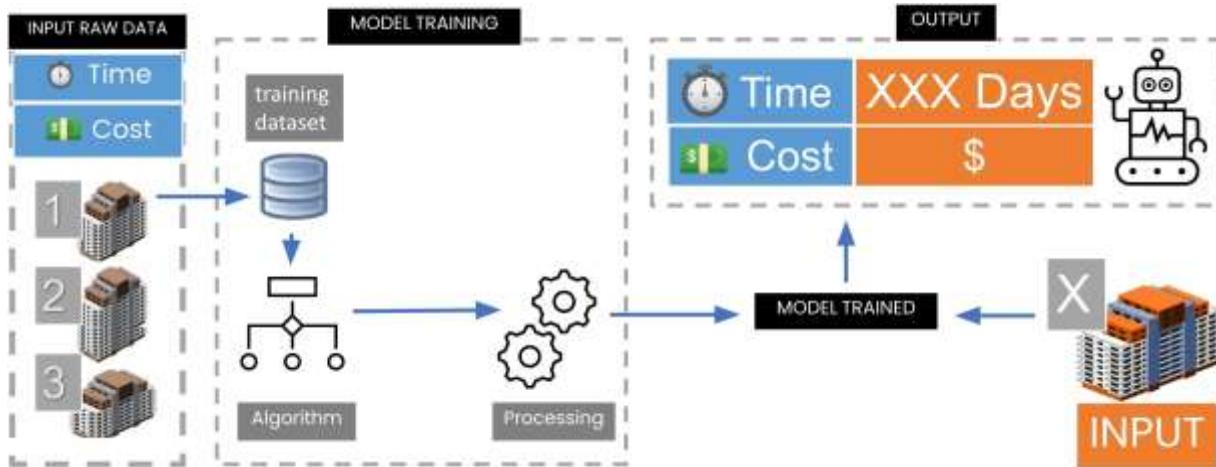
Strojové učenie eliminuje manuálne vyhľadávanie vzorcov a nahradza ich vyškolenými algoritmami, ktoré nezávisle identifikujú závislosti, ktoré sú mnohonásobne presnejšie ako akékoľvek vopred stanovené rovnice. Prípadne vytvorme algoritmus strojového učenia, ktorý vytvorí model na základe predchádzajúceho pochopenia problému a historických údajov, ktoré môžu byť neúplné (obrázok 9.2-14).

Na príklade cien strojové učenie umožňuje vytvárať rôzne typy matematických modelov, ktoré nevyžadujú znalosť presného mechanizmu tvorby nákladov. Model sa "učí" z údajov o predchádzajúcich projektoch, pričom sa prispôsobuje reálnym zákonitostiam medzi parametrami stavby, jej nákladmi a termínmi.



Obr. 9.2-14 Na rozdiel od klasického odhadu založeného na vzorcoch je algoritmus strojového učenia trénovaný na historických údajoch.

V kontexte strojového učenia pod dohľadom, každý projekt v súbore trénovaných údajov obsahuje vstupné atribúty (napr. náklady a čas podobných budov) a očakávané výstupné hodnoty (napr. náklady alebo čas). Podobný súbor údajov sa používa na vytvorenie a prispôsobenie modelu strojového učenia (obrázok 9.2-15). Čím väčší je súbor údajov a čím vyššia je kvalita údajov v ňom, tým presnejší bude model a tým presnejšie budú výsledky predpovedí.



Obrázok 9.2-15 ML model natrénonaný na údajoch o nákladoch a harmonograme z minulých projektov určí náklady a harmonogram nového projektu s určitou pravdepodobnosťou.

Po vytvorení modelu a jeho vyškolení na odhad výstavby nového projektu stačí modelu poskytnúť nové atribúty nového projektu a model poskytne odhadované výsledky na základe predtým naučených vzorov s určitou pravdepodobnosťou.

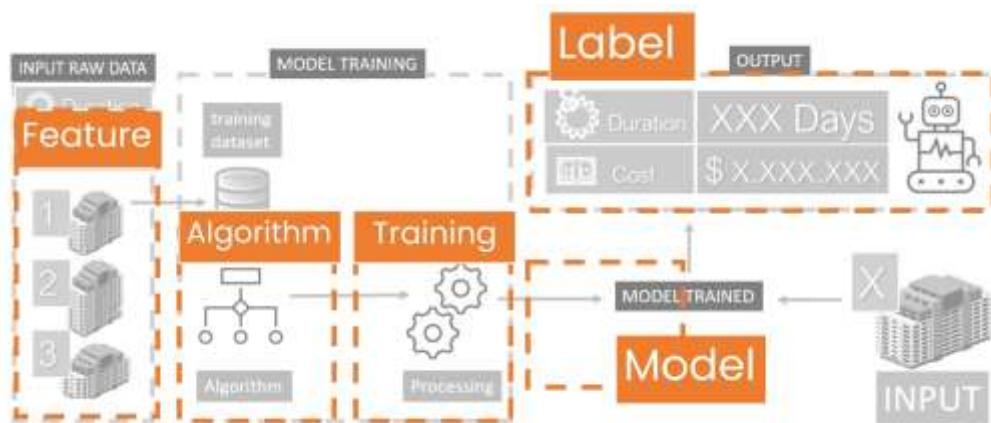
## Kľúčové koncepty strojového učenia

Strojové učenie nie je mágia, je to len matematika, údaje a hľadanie vzorov. Nemá žiadnu skutočnú inteligenciu, ale je to program vycvičený na základe údajov, ktorý rozpoznáva vzory a prijíma

rozhodnutia bez neustálej účasti človeka.

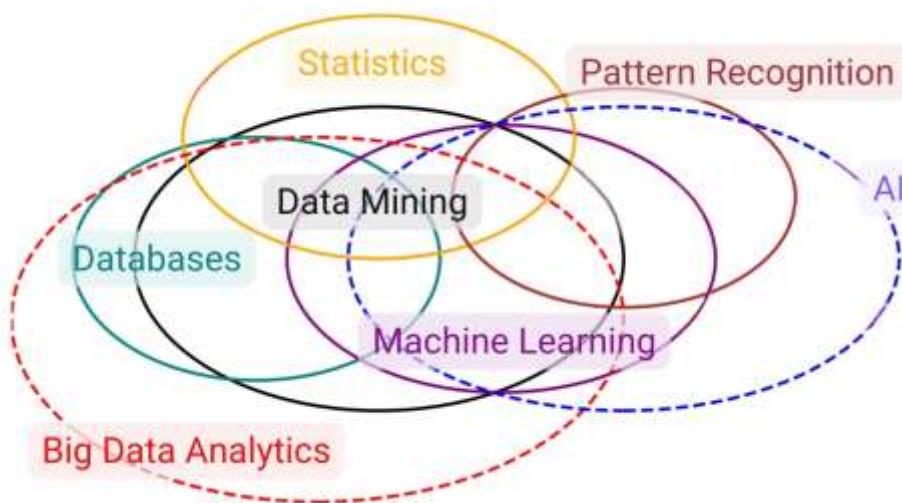
Strojové učenie používa na opis svojej štruktúry niekoľko kľúčových pojmov (obrázok 9.2-16):

- **Štítky** sú cieľové premenné alebo atribúty (parameter "Survivor" v súbore údajov Titanic), ktoré by mal model predpovedať. Príklad: náklady na výstavbu (napr. v dolároch), trvanie stavebných prác (napr. v mesiacoch).
- **Funkcie** sú nezávislé premenné alebo atribúty, ktoré slúžia ako vstupy do modelu. V prognostickom modeli sa používajú na predpovedanie značiek. Príklady: veľkosť pozemku (v metroch štvorcových), počet poschodí budovy, celková podlahová plocha budovy (v metroch štvorcových), geografická poloha (zemepisná šírka a dĺžka), typ materiálov použitých pri výstavbe. Počet charakteristík určuje aj rozmernosť údajov.
- **Model** je súbor rôznych hypotéz, z ktorých jedna aproximuje cieľovú funkciu, ktorá sa má predpovedať alebo approximovať. Príklad: model strojového učenia, ktorý využíva techniky regresnej analýzy na predpovedanie nákladov a termínov výstavby.
- **Algoritmus učenia** Algoritmus učenia je proces hľadania najlepšej hypotézy v modeli, ktorá presne zodpovedá cieľovej funkcií, pomocou množiny trénovalých údajov. Príklad: Algoritmus lineárnej regresie, KNN alebo náhodného lesa, ktorý analyzuje údaje o nákladoch a čase výstavby s cieľom identifikovať vzťahy a vzory.
- **Trénovanie** - počas procesu trénovalia algoritmus analyzuje trénovalé údaje a nachádza vzory, ktoré zodpovedajú vzťahu medzi vstupnými atribútmi a cieľovými značkami. Výsledkom tohto procesu je natrénovaný model strojového učenia, pripravený na predikciu. Príklad: Proces, v ktorom algoritmus analyzuje historické údaje o stavbe (náklady, čas, charakteristiky objektu) s cieľom vytvoriť predikčný model.



Obrázok 9.2-16 ML používa značky a atribúty na vytvorenie modelov, ktoré sú vyškolené na údajoch pomocou algoritmov na predpovedanie výsledkov.

Strojové učenie neexistuje izolované, ale je súčasťou širšieho ekosystému analytickej disciplíny vrátane štatistiky, databáz, dolovania údajov, rozpoznávania vzorov, analýzy veľkých objemov údajov a umelej inteligencie. Na obrázku 9.2-17 je znázorené, ako sa tieto oblasti prekrývajú a dopĺňajú, čím poskytujú komplexný rámec pre moderné rozhodovacie a automatizačné systémy.



Obr. 9.2-17 Vzťahy medzi jednotlivými oblastami analýzy údajov: štatistika, strojové učenie, umelá inteligencia, big data, rozpoznávanie vzorov a dolovanie údajov.

Hlavným cieľom strojového učenia je vybaviť počítače schopnosťou automaticky sa učiť znalosti bez ľudského zásahu alebo pomoci a podľa toho prispôsobiť svoje činnosti [159].

V budúcnosti bude teda úlohou človeka iba poskytnúť stroju kognitívne schopnosti - nastaví podmienky, váhy a parametre a o zvyšok sa postará model strojového učenia.

V ďalšej kapitole sa budeme zaoberať konkrétnymi príkladmi použitia algoritmov. Budú použité reálne tabuľky a zjednodušené modely, aby sme ukázali, ako sa prognóza vytvára krok za krokom.



## KAPITOLA 9.3.

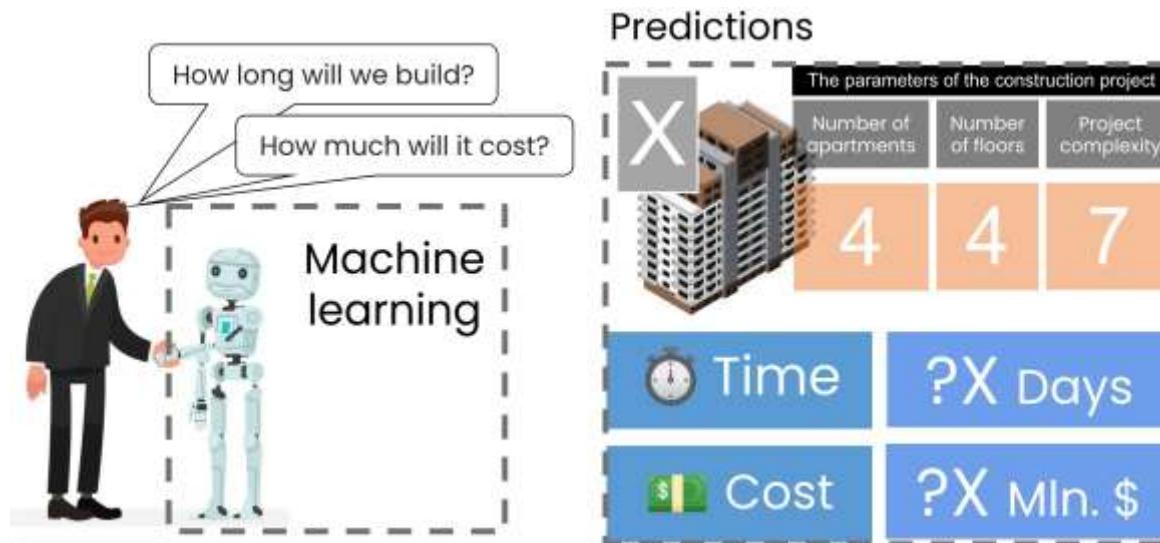
### PREDPOVEDANIE NÁKLADOV A TERMÍNOV POMOCOU STROJOVÉHO UČENIA

#### Príklad použitia strojového učenia na zistenie nákladov a časového harmonogramu projektu

Odhad času výstavby a nákladov je jedným z kľúčových procesov v činnosti stavebnej spoločnosti. Tradične takéto odhady vykonávajú odborníci na základe skúseností, referenčných kníh a regulačných databáz. S digitálnou transformáciou a rastúcou dostupnosťou údajov je však v súčasnosti možné využívať modely strojového učenia (ML) na zvýšenie presnosti a automatizáciu takýchto odhadov.

Zavedenie strojového učenia do procesu výpočtu nákladov a termínov výstavby umožňuje nielen efektívnejšie plánovanie, ale stáva sa aj východiskom pre integráciu inteligentných modelov do ďalších podnikových procesov - od riadenia rizík až po optimalizáciu logistiky a obstarávania.

Je dôležité vedieť rýchlo určiť, ako dlho bude trvať realizácia projektu a aké budú jeho celkové náklady. Tieto otázky týkajúce sa času a nákladov na projekt sú tradične v popredí záujmu klientov aj stavebných spoločností od vzniku stavebníctva.



Obr. 9.3-1 V stavebných projektoch sú rýchlosť a kvalita odhadu času výstavby a nákladov kľúčovými faktormi úspechu.

V nasledujúcom príklade sa získajú kľúčové údaje z minulých projektov a použijú sa na vytvorenie modelu strojového učenia, ktorý nám umožní použiť model na odhad nákladov a časového harmonogramu nových stavebných projektov s novými parametrami (obrázok 9.3-1).

Uvažujte o troch projektoch s troma kľúčovými atribútmi: počtom bytov (kde 100 bytov zodpovedá číslu 10 kvôli jednoduchšej vizualizácii), počtom poschodí a podmieneným meradlom zložitosti stavby na

stupnici od 1 do 10, kde 10 je najvyššie skóre zložitosti. V strojovom učení sa proces prevodu a zjednodušenia hodnôt, ako je 100 na 10 alebo 50 na 5, nazýva "normalizácia".

**Normalizácia** v strojovom učení je proces zosúladenia rôznych číselných údajov na spoločnú stupnicu s cieľom uľahčiť ich spracovanie a analýzu. Tento proces je dôležitý najmä vtedy, keď majú údaje rôzne stupnice a merné jednotky.

Predpokladajme, že prvý projekt (obrázok 9.3-2) mal 50 bytov (po normalizácii 5), 7 poschodí a skóre zložitosti 2, čo znamenalo relatívne jednoduchú stavbu. Druhý projekt už mal 80 bytov, 9 poschodí a relatívne zložitú konštrukciu. Za týchto podmienok trvala výstavba prvého a druhého bytového domu 270 a 330 dní a celkové náklady na projekt boli 4,5 milióna USD a 5,8 milióna USD.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment	Number of floors	Project complexity	Time	Cost
1	5	7	2	270	\$ 4.502.000
2	8	9	6	330	\$ 5.750.000
3	3	5	3	230	\$ 3.262.000
X	4	4	7	?X	\$ ?X. XXX.XXX

Obrazok 9.3-2 Príklad súboru minulých projektov, ktorý sa použije na odhad času a nákladov budúceho projektu X.

Pri vytváraní modelu strojového učenia pre takéto údaje je hlavnou úlohou identifikovať kritické atribúty (alebo značky) na predikciu, v tomto prípade čas výstavby a náklady. Pri malom súbore údajov použijeme informácie o predchádzajúcich stavebných projektoch na plánovanie nových projektov: pomocou algoritmov strojového učenia, musíme predpovedať náklady a trvanie výstavby nového projektu X na základe daných atribútov nového projektu, ako je 40 bytov, 4 poschodia a relatívne vysoká zložitosť projektu 7 (obr. 9.3-2). V reálnom prostredí môže byť počet vstupných parametrov oveľa väčší, od niekoľkých desiatok až po stovky faktorov. Môžu to byť napríklad: typ stavebných materiálov, klimatické pásmo, kvalifikačná úroveň dodávateľov, dostupnosť inžinierskych sietí, typ založenia, sezóna začatia prác, pripomienky majstrov atď.

Na vytvorenie prediktívneho modelu strojového učenia, musíme vybrať algoritmus na jeho vytvorenie. Algoritmus v strojovom učení je ako matematický recept, ktorý učí počítač, ako robiť predpovede (miešať v správnom poradí parametre) alebo prijímať rozhodnutia na základe údajov.

Na analýzu údajov o minulých stavebných projektoch a predpovedanie času a nákladov budúcich projektov (obrázok 9.3-2) možno použiť jeden z populárnych algoritmov strojového učenia:

- **Lineárna regresia (Lineárna regresia):** tento algoritmus sa snaží nájsť priamy vzťah medzi atribútmi, napríklad medzi počtom poschôd a stavebnými nákladmi. Cieľom algoritmu je nájsť lineárnu rovnicu, ktorá najlepšie opisuje tento vzťah, čo umožňuje robiť predpovede.
- **Algoritmus k-najbližších susedov (k-NN):** tento algoritmus porovnáva nový projekt s predchádzajúcimi projektmi, ktoré boli podobné veľkosťou alebo zložitosťou. K-NN klasifikuje údaje na základe toho, ktoré z k (počtu) trénovaných príkladov sú mu najbližšie. V kontexte regresie je výsledkom priemer alebo medián k najbližších susedov.
- **Rozhodovacie stromy:** je model prediktívneho modelovania, ktorý rozdeľuje údaje do podmnožín na základe rôznych podmienok pomocou stromovej štruktúry. Každý uzol stromu predstavuje podmienku alebo otázku vedúcu k ďalšiemu rozdeleniu údajov a každý list predstavuje konečnú predpoveď alebo výsledok. Algoritmus rozdeľuje údaje do menších skupín na základe rôznych charakteristik, napr. najprv podľa počtu príbehov, potom podľa zložitosti a tak ďalej, aby bolo možné vytvoriť predpoved.

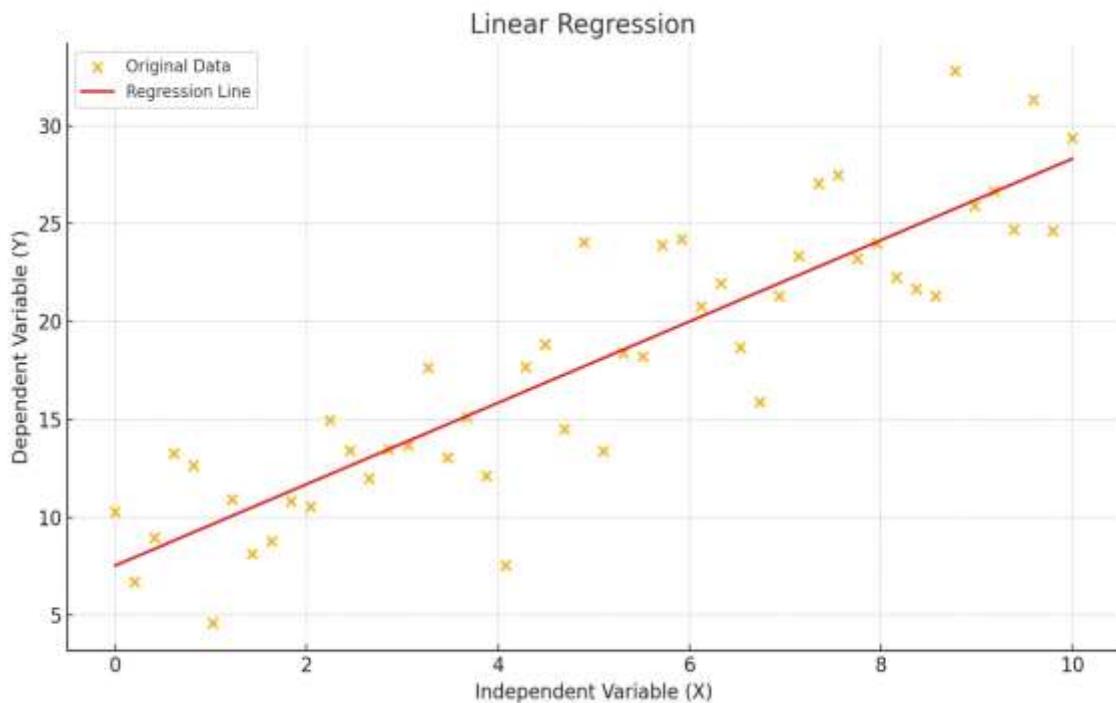
Pozrime sa na algoritmy strojového učenia na odhadovanie nákladov nového projektu na príklade dvoch populárnych algoritmov: lineárnej regresie a algoritmu K-najbližších susedov.

## Predpovedanie nákladov a času projektu pomocou lineárnej regresie

Lineárna regresia je základný algoritmus analýzy údajov, ktorý predpovedá hodnotu premennej na základe lineárneho vzťahu s jednou alebo viacerými inými premennými. Tento model predpokladá, že medzi závislou premenou a jednou alebo viacerými nezávislými premennými existuje priamy lineárny vzťah, a cieľom algoritmu je tento vzťah nájsť.

Jednoduchosť a prehľadnosť lineárnej regresie z nej urobila populárny nástroj v rôznych oblastiach. Pri práci s jednou premenou je lineárna regresia o hľadaní najlepšie pasujúcej priamky cez dátové body.

Lineárna regresia nájde najlepšiu priamku (červená čiara), ktorá aproximuje závislosť medzi vstupnou premenou X a výstupnou premenou Y. Táto priamka umožňuje predpovedať hodnoty Y pre nové hodnoty X na základe identifikovaného lineárneho vzťahu (obrázok 9.3-3).



Obr. 9.3-3 Princíp lineárnej regresie spočíva v hľadaní najlepšej priamky cez trénované hodnoty.

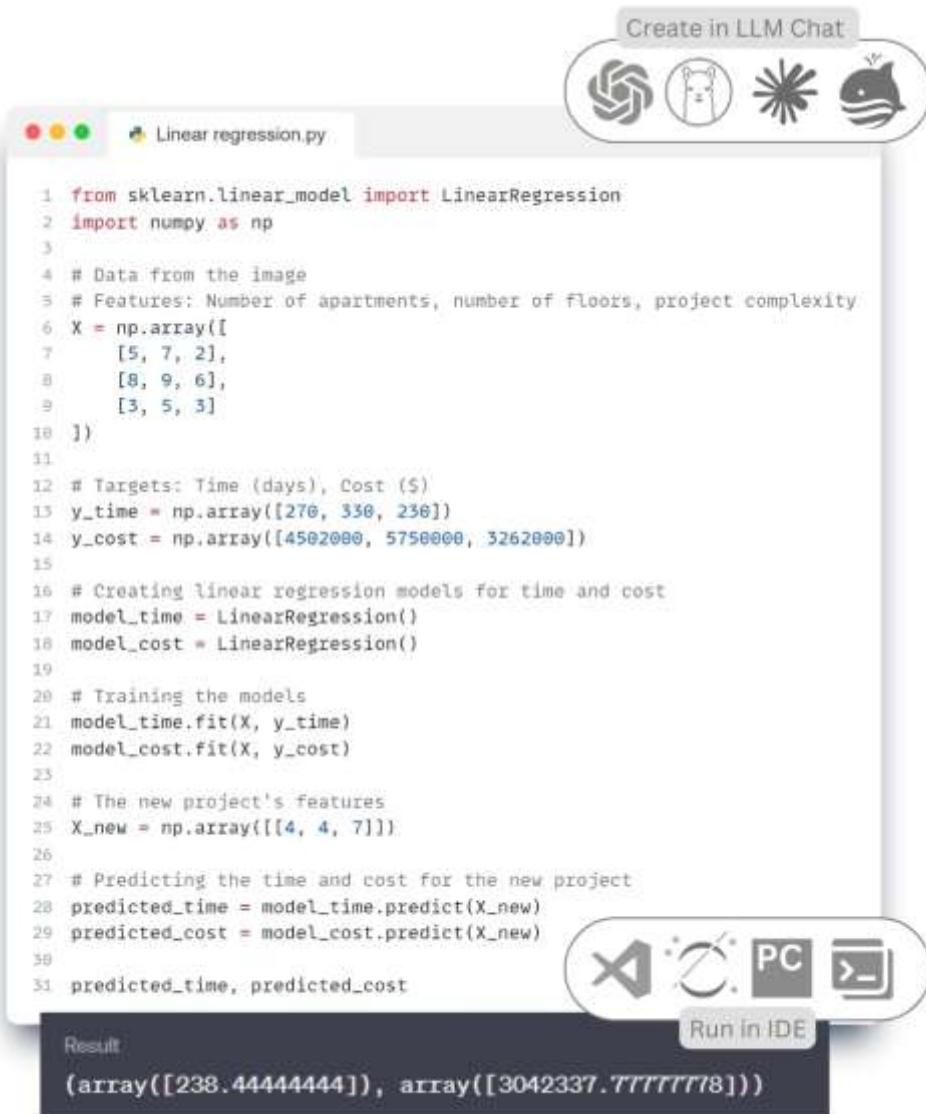
Táto čiara je reprezentovaná rovnicou, v ktorej zadáním hodnoty nezávislej premennej (X) vznikne predpokladaná hodnota závislej premennej (Y). Tento proces umožňuje efektívne predpovedať Y zo známych hodnôt X pomocou lineárneho vzťahu medzi nimi. Príklad nájdienia takejto štatisticky spriemerovanej priamky možno vidieť v hodnotení údajov o stavebných povoleniach v San Franciscu (obrázok 9.1-7), kde bola vypočítaná inflácia pomocou lineárnej regresie pre rôzne typy zariadení.

Načítajme tabuľku s údajmi o projekte (obrázok 9.3-2 z predchádzajúcej kapitoly) priamo do LLM a požiadajme ho, aby pre nás vytvoril jednoduchý model strojového učenia.

- ❷ Pošlite textovú požiadavku do chatu LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Potrebuje ukázať konštrukciu jednoduchého modelu strojového učenia na predpovedanie nákladov a času na realizáciu nového projektu X (obr. 9.3-2 ako priložený obrázok) ↴

💡 Odpoveď LLM :



The screenshot shows a software interface for developing machine learning models. At the top, there's a toolbar with icons for creating in LLM Chat, a neural network, a cat, a flower, and a whale. Below that is a window titled "Linear regression.py" containing the following Python code:

```

1 from sklearn.linear_model import LinearRegression
2 import numpy as np
3
4 # Data from the image
5 # Features: Number of apartments, number of floors, project complexity
6 X = np.array([
7     [5, 7, 2],
8     [8, 9, 6],
9     [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

At the bottom, there's a "Result" section showing the output of the code execution:

```
(array([238.44444444]), array([3042337.77777778]))
```

Below the result is a "Run in IDE" button.

Obrázok 9.3-4 LLM si vybral lineárnu regresiu na vytvorenie modelu strojového učenia na predpovedanie nákladov a času projektu.

Program LLM automaticky rozpoznal tabuľku z priloženého obrázka a previedol údaje z vizuálneho formátu na tabuľkové pole (obr. 9.3-4 - riadok 6). Toto pole sa použilo ako základ na vytvorenie príznakov a značiek, z ktorých sa vytvoril model strojového učenia (obr. 9.3-4 - 17. - 22. riadok), ktorý využíval lineárnu regresiu.

Pomocou základného lineárneho regresného modelu, ktorý bol vycvičený na "extrémne malom" súbore údajov, sa vykonali predpovede pre nový hypotetický stavebný projekt označený ako projekt X. V

našom probléme je tento projekt charakterizovaný 40 bytmi, 4 poschodiami a úrovňou zložitosti 7 (obrázok 9.3-2).

Ako predpovedá lineárny regresný model založený na obmedzenom a malom súbore údajov pre nový projekt X (obrázok 9.3-4 - riadok 24-29):

- **Výstavba bude trvať** približne 238 dní (238.444444)
- **Celkové výdavky** budú predstavovať približne \$ 3 042 338 (3042337,777)

Na ďalšie skúmanie hypotézy o nákladoch projektu je užitočné experimentovať s rôznymi algoritmami a metódami strojového učenia. Preto predpovedajme rovnaké hodnoty nákladov a času pre nový projekt X na základe malého súboru historických údajov pomocou algoritmu K-Nearest Neighbours (k-NN).

### **Predpovede nákladov a času projektu pomocou algoritmu K-najbližší sused (k-NN)**

Ako ďalší prediktor na odhad nákladov a trvania nového projektu používame algoritmus k-Nearest Neighbours (k-NN). Algoritmus k-Nearest Neighbours (k-NN) je metóda strojového učenia pod dohľadom (supervised machine learning) na klasifikáciu aj regresiu. O algoritme k-NN sme už hovorili aj v súvislosti s vyhľadávaním vo vektorovej databáze (obrázok 8.2-2), kde sa používa na vyhľadávanie najbližších vektorov (napr. textov, obrázkov alebo technických popisov). V tomto prístupe je každý projekt reprezentovaný ako bod vo viacrozmernom priestore, kde každá dimenzia zodpovedá konkrétnemu atribútu projektu.

V našom prípade, vzhľadom na tri atribúty každého projektu, ich budeme reprezentovať ako body v trojrozmernom priestore (obrázok 9.3-5). Náš pripravovaný projekt X bude teda lokalizovaný v tomto priestore so súradnicami ( $x=4$ ,  $y=4$ ,  $z=7$ ). Treba poznamenať, že v reálnych podmienkach môže byť počet bodov a dimenzionalita priestoru rádovo väčšia.

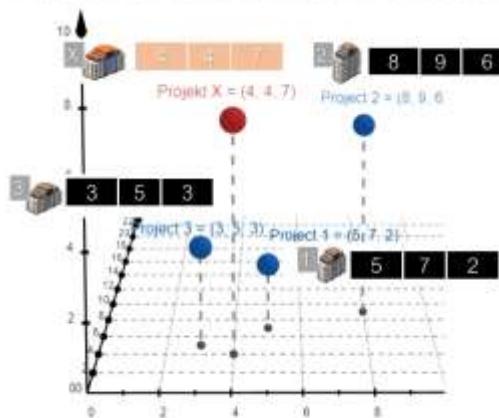
Algoritmus K-NN (k-nearest neighbours) funguje na základe merania vzdialenosť medzi požadovaným projektom X a projektmi v trénovacej databáze. Porovnaním týchto vzdialenosťí algoritmus určí projekty, ktoré sú najbližšie k bodu nového projektu X.

Ak je napríklad druhý projekt ( $x=8$ ,  $y=9$ ,  $z=6$ ) z nášho pôvodného súboru údajov oveľa ďalej od X (obrázok 9.3-5) ako ostatné projekty, možno ho z ďalšej analýzy vylúčiť. V dôsledku toho možno na výpočty použiť len dva ( $k=2$ ) najbližšie projekty, na základe ktorých sa určí priemerná hodnota.

Takáto metóda prostredníctvom vyhľadávania v susedstve umožňuje posúdiť podobnosti medzi projektmi, čo následne pomáha vyvodiť závery o pravdepodobných nákladoch a načasovaní nového projektu na základe podobných projektov, ktoré boli realizované v minulosti.

### k-nearest neighbors algorithm

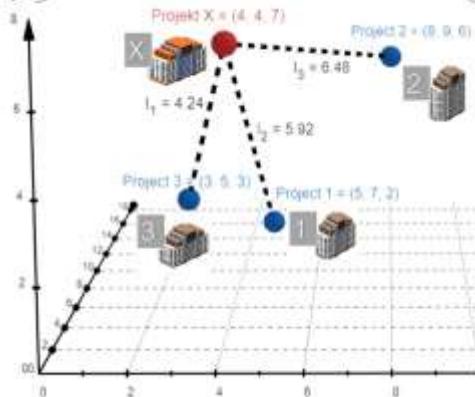
The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



Distance between two  

$$\text{length} = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$$

5,92 ✓	
6,48 ✗	
4,24 ✓	



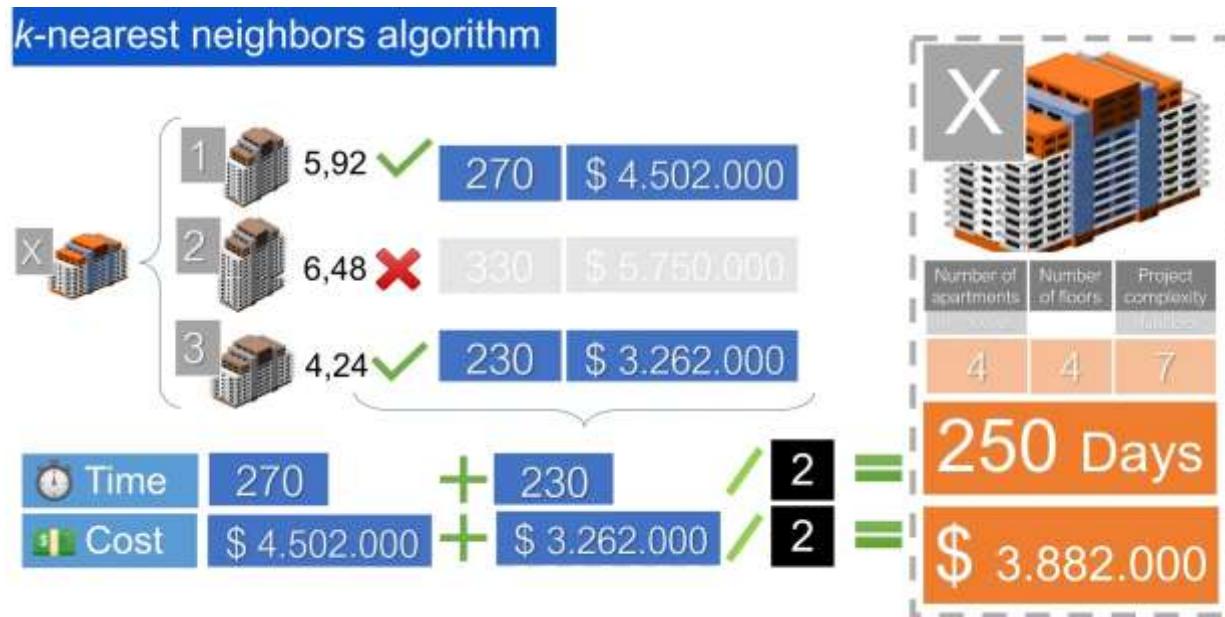
Obr. 9.3-5 V algoritme K-NN sú projekty reprezentované ako body vo viacozmernom priestore a na základe vzdialenosí sa vyberajú najbližšie projekty na posúdenie podobnosti a predikciu.

Práca s k-NN zahŕňa niekoľko kľúčových krokov:

- **Príprava údajov:** najprv sa načítajú tréningové a testovacie súbory údajov. Tréningové údaje sa používajú na "trénovanie" algoritmu a testovacie údaje sa používajú na kontrolu jeho účinnosti.
- **Výber parametra K:** vyberie sa číslo K, ktoré udáva, kol'ko najbližších susedov (dátových bodov) sa má v algoritme zohľadniť. Hodnota "K" je veľmi dôležitá, pretože ovplyvňuje výsledok.
- Proces klasifikácie a regresia pre testovacie údaje:
  - **Výpočet vzdialenosí:** pre každý prvok z testovacích údajov sa vypočíta vzdialenosť ku každému prvku z tréningových údajov (obrázok 9.3-5). Na tento účel sa môžu použiť rôzne metódy merania vzdialenosí, napríklad Euklidova vzdialenosť (najbežnejšia metóda), Manhattanova vzdialenosť alebo Hammingova vzdialenosť.
  - **Triedenie a výber K najbližších susedov:** po výpočte vzdialenosí sa tieto zoradia a vyberie sa K najbližších bodov k testovanému bodu.
  - **Určenie triedy alebo hodnoty testovacieho bodu:** ak ide o klasifikačnú úlohu, trieda testovacieho bodu sa určí na základe najčastejšej triedy spomedzi K vybraných susedov. Ak ide o regresnú úlohu, vypočíta sa priemer (alebo iná miera centrálnej tendencie) hodnôt K susedov.
- **Ukončenie procesu:** po klasifikácii všetkých testovacích údajov alebo po vykonaní predpovedí pre ne je proces ukončený.

Algoritmus k-najbližších susedov (k-NN) je účinný v mnohých praktických aplikáciách a je jedným z hlavných nástrojov v arzenáli odborníkov na strojové učenie. Tento algoritmus je oblúbený vďaka svojej jednoduchosti a efektívnosti, najmä v úlohách, kde sa vzťahy medzi údajmi dajú ľahko interpretovať.

V našom príklade boli po použití algoritmu K-najbližších susedov identifikované dva projekty (z našej malej vzorky) s najkratšou vzdialenosťou od projektu X (obrázok 9.3-5). Na základe týchto projektov algoritmus určí priemer ich ceny a dĺžky výstavby. Po analýze (obr. 9.3-6) algoritmus na základe spriemerovania najbližších susedov dospel k záveru, že projekt X bude stáť približne \$ 3 800 000 USD a jeho realizácia potrvá približne 250 dní.



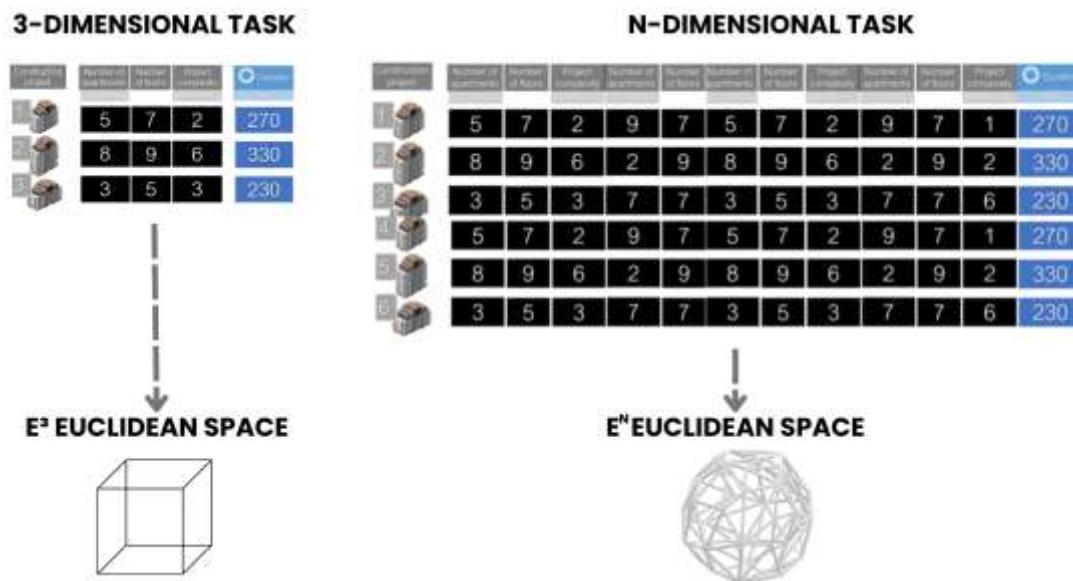
Obrázok 9.3-6 Algoritmus K-najbližších susedov určuje náklady a harmonogram projektu X analýzou dvoch najbližších projektov vo vzorke.

Algoritmus k-Nearest Neighbors (k-NN) je obzvlášť populárny v úlohách klasifikácie a regresie, ako sú napríklad odporúčacie systémy, kde sa používa na navrhovanie produktov alebo obsahu na základe preferencií podobných záujmom konkrétnego používateľa. Okrem toho sa k-NN široko používa v lekárskej diagnostike na klasifikáciu typov chorôb na základe symptómov pacienta, pri rozpoznávaní vzorov a vo finančnom sektore na posudzovanie bonity zákazníkov.

Modely strojového učenia môžu aj s obmedzenými údajmi poskytovať užitočné predpovede a výrazne zlepšiť analytickú zložku riadenia stavebných projektov. S rozširovaním a čistením historických údajov je možné prejsť na sofistikovanejšie modely - napríklad zohľadňujúce typ stavby, lokalitu, sezónu začiatku výstavby a ďalšie faktory.

V našej zjednodušenej úlohe sme na vizualizáciu v 3D priestore použili tri atribúty, ale skutočné

projekty v priemere obsahujú stovky alebo tisíce atribútov (pozri súbor údajov z kapitoly "Príklad veľkých dát na báze CAD (BIM)", čo výrazne zvyšuje rozmernosť priestoru a zložitosť reprezentácie projektov ako vektorov (obrázok 9.3-7).



Obrázok 9.3-7 V zjednodušenom príklade boli pre 3D -vizualizáciu použité tri atribúty, zatiaľ čo skutočné projekty ich majú viac.

Použitím rôznych algoritmov na rovnaký súbor údajov pre projekt X, ktorý má 40 bytov, 4 poschodia a úroveň zložitosti 7, sa získali rôzne predpovedané hodnoty. Algoritmus lineárnej regresie predpovedal čas dokončenia 238 dní a náklady \$ 3 042 338 (obrázok 9.3-4), zatiaľ čo algoritmus k-NN predpovedal 250 dní a \$ 3 882 000 (obrázok 9.3-6).

Presnosť predpovedí získaných pomocou modelov strojového učenia, priamo závisí od objemu a kvality vstupných údajov. Čím viac projektov je zapojených do tréningu a čím úplnejšie a presnejšie sú reprezentované ich charakteristiky (atribúty) a výsledky (značky), tým vyššia je pravdepodobnosť získania spoločlivých predpovedí s minimálnymi hodnotami chýb.

V tomto procese zohrávajú dôležitú úlohu techniky predbežného spracovania údajov vrátane:

- Normalizácia na zosúladenie prvkov do spoločnej stupnice;
- Detekcia a eliminácia odchýlok, odstránenie skreslenia modelu;
- Kódovanie kategorických atribútov, ktoré umožňuje manipuláciu s textovými údajmi;
- Doplnenie chýbajúcich hodnôt, čím sa zvýší robustnosť modelu.

Okrem toho sa na posúdenie všeobecnosti modelu a jeho odolnosti voči novým súborom údajov používajú metódy krízového overovania s cieľom odhaliť nadmerné prispôsobenie a zvýšiť spoločlivosť predpovede.

*Chaos je poriadok, ktorý treba rozlúštiť [160].*

- José Saramago, "Dvojník"

Aj keď sa vám zdá, že chaos vašich úloh sa nedá formálne opísať, mali by ste vedieť, že každá udalosť vo svete a najmä stavebné procesy podliehajú matematickým zákonitostiam, ktoré môžu potrebovať podporu výpočtu hodnôt nie prostredníctvom striktných vzorcov, ale s pomocou štatistiky a historických údajov.

Tradičné odhady oddelenia kalkulácie nákladov aj modely strojového učenia nevyhnutne čelia neistote a potenciálnym zdrojom chýb. Ak je však k dispozícii dostatok kvalitných údajov, modely strojového učenia môžu vykazovať porovnatelnú a niekedy dokonca vyššiu presnosť predpovedí ako expertné odhady.

Strojové učenie sa pravdepodobne stane spoľahlivým doplnkovým nástrojom pre analýzy, ktoré môžu: spresniť výpočty, navrhnuť alternatívne scenáre a identifikovať skryté závislosti medzi parametrami projektu. Takéto modely si nebudú nárokovala na univerzálnosť, ale čoskoro zaujmú dôležité miesto vo výpočtoch a rozhodovacích procesoch. Technológie strojového učenia nevylúčia účasť inžinierov, odhadcov a analytikov, ale naopak, rozšíria ich možnosti tým, že ponúknu ďalší pohľad založený na historických údajoch.

Ak sa strojové učenie správne integruje do obchodných procesov stavebných spoločností, má potenciál stať sa dôležitým prvkom v systéme podpory rozhodovania manažmentu - nie ako náhrada človeka, ale ako rozšírenie jeho profesionálnej intuície a inžinierskej logiky.

## Ďalšie kroky: od skladovania k analýze a prognózovaniu

Moderné prístupy k práci s údajmi začínajú meniť spôsob rozhodovania v stavebníctve. Prechod od intuitívneho hodnotenia k objektívnej analýze údajov nielenže zvyšuje presnosť, ale otvára aj nové možnosti optimalizácie procesov. Na zhrnutie tejto časti je vhodné zdôrazniť hlavné praktické kroky, ktoré vám pomôžu aplikovať diskutované metódy pri vašich každodenných úlohách:

- Vytvorenie udržateľnej skladovacej infraštruktúry
- Pokúste sa spojiť rôznorodé dokumenty a údaje o projekte do jedného tabuľkového modelu, pričom kľúčové informácie sústredite do jedného dátového rámca na ďalšiu analýzu.
- Používajte efektívne formáty ukladania údajov - napríklad stĺpcové formáty ako Apache Parquet namiesto CSV alebo XLSX - najmä pre súbory, ktoré by sa v budúcnosti mohli použiť na trénovanie modelov strojového učenia.
- Zavedenie systému verzií údajov na sledovanie zmien v priebehu projektu

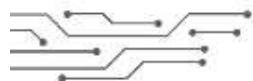
**■ Implementácia analytických a automatizačných nástrojov**

- Začnite analyzovať historické údaje o projektoch - podľa dokumentácie, modelov, odhadov - s cieľom identifikovať vzory, trendy a anomálie.
- Hlavné procesy ETL (Extract, Transform, Load) na automatické načítanie a prípravu údajov.
- Naučte sa vizualizovať klúčové metriky pomocou rôznych bezplatných vizualizačných knižníc Pythonu
- začať uplatňovať štatistické metódy a náhodný výber vzoriek na získanie reprezentatívnych a reprodukovateľných analytických zistení

**■ Zvyšovanie vyspelosti pri práci s údajmi**

- Naučte sa niekoľko základných algoritmov strojového učenia na jednoduchých a priamych príkladoch, ako je súbor údajov o Titaniku
- Analyzujte súčasné procesy a zistite, kde môžete prejsť od rigidnej logiky príčin a následkov k štatistickým metódam prognózovania a odhadovania.
- Začnite sa k údajom správať ako k strategickému aktívu, a nie ako k vedľajšiemu produktu: rozhodovacie procesy budujte skôr na základe dátových modelov než na základe konkrétnych softvérových riešení.

Stavebné spoločnosti, ktoré si uvedomili hodnotu údajov, vstupujú do novej fázy rozvoja, v ktorej konkurenčnú výhodu neurčuje množstvo zdrojov, ale rýchlosť rozhodovania na základe analytických údajov.



## MAXIMÁLNE POHODLIE S TLAČENOU VERZIOU

Držíte v rukách bezplatnú digitálnu verziu knihy **Data-Driven Construction**. Pre pohodlnejšiu prácu a rýchly prístup k materiálom vám odporúčame venovať pozornosť **tlačenému vydaniu**:



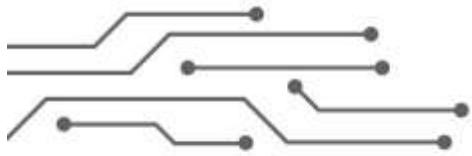
■ **Vždy po ruke:** kniha v tlačenej podobe sa stane spoľahlivým pracovným nástrojom, ktorý vám umožní rýchlo nájsť a použiť správne vizualizácie a diagramy v akejkoľvek pracovnej situácii

■ **Vysoká kvalita ilustrácií:** všetky obrázky a grafika v tlačenom vydani sú prezentované v maximálnej kvalite

■ **Rýchly prístup k informáciám:** jednoduchá navigácia, možnosť robiť si poznámky, záložky a pracovať s knihou kdekoľvek.

Zakúpením plnej tlačenej verzie knihy získate praktický nástroj na pohodlnú a efektívnu prácu s informáciami: možnosť rýchlo využívať vizuálne materiály pri každodenných úlohách, rýchlo nájsť potrebné schémy a robiť si poznámky. Okrem toho svojou kúpou podporíte šírenie otvorených poznatkov.

Tlačenú verziu knihy si môžete objednať na adrese: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## X ČASŤ

# **STAVEBNÍCTVO V DIGITÁLNOM VEKU. PRÍLEŽITOSTI A VÝZVY**

Posledná desiaťa časť je komplexným pohľadom na budúlosť stavebníctva v ére digitálnej transformácie. Analyzuje prechod od kauzánej analýzy k práci s koreláciami veľkých dát. Uvádzajú sa paralely medzi vývojom výtvarného umenia a vývojom práce s dátami v stavebníctve, pričom sa ukazuje, ako sa odvetvie posúva od detailnej kontroly k holistikému chápaniu procesov.

Skúma sa koncepcia "uberizácie" stavebníctva, kde transparentnosť údajov a automatizované výpočty môžu radikálne zmeniť tradičné obchodné modely, odstrániť potrebu sprostredkovateľov a znížiť možnosti špekulácií. Podrobne sa rozoberajú nevyriešené otázky, ako napríklad univerzálna klasifikácia prvkov, čo dáva stavebným spoločnostiam čas na prispôsobenie sa novému prostrediu. Časť sa uzatvára konkrétnymi odporúčaniami na formovanie stratégie digitálnej transformácie, ktorá zahŕňa analýzu slabých miest a rozšírenie služieb s cieľom udržať si konkurencieschopnosť v meniacom sa odvetví.

## KAPITOLA 10.1.

### STRATÉGIE PREŽITIA: BUDOVANIE KONKURENČNÝCH VÝHOD

#### Korelácie namiesto výpočtov: budúcnosť stavebnej analytiky

V dôsledku rýchlej digitalizácie informácií (obrázok 1.1-5) prechádza moderné stavebníctvo zásadnou transformáciou, v rámci ktorej sa údaje stávajú nielen nástrojom, ale strategickým aktívom, ktoré môže zásadne zmeniť tradičné prístupy k riadeniu projektov a podnikov.

Tisíce rokov sa stavebné činnosti spoliehali na deterministické metódy - presné výpočty, vypracovanie detailov a prísnu kontrolu parametrov. V prvých storočiach nášho letopočtu rímski inžinieri uplatňovali matematické princípy pri stavbe akvaduktov a mostov. V stredoveku sa architekti usilovali o ideálne proporcie gotických katedrál a v 20. storočí sa v rámci industrializácie vytvorili systémy štandardizovaných noriem a predpisov, ktoré sa stali základom pre hromadnú výstavbu.

V súčasnosti sa vektor vývoja posúva od hľadania striktne príčinných vzťahov k pravdepodobnostnej analýze, hľadaniu korelácií a skrytých vzorov. Odvetvie vstupuje do novej fázy - klíčovým zdrojom sa stávajú údaje a na nich založená analytika nahradza intuitívne a lokálne optimalizované prístupy.



Obr. 10.1-1 Skrytý potenciál konštrukčných údajov: existujúce výpočty v podniku sú len špičkou ľadovca, ktorý má manažment k dispozícii na analýzu.

Informačný systém spoločnosti je ako ľadovec (obr. 10.1-1): manažment spoločnosti vidí len malú časť potenciálu údajov, zatiaľ čo hlavná hodnota je skrytá v hlbínach. Dôležité je hodnotiť údaje nielen z hľadiska ich súčasného využitia, ale aj z hľadiska príležitostí, ktoré sa vďaka nim uvoľnia v budúcnosti. Práve tie spoločnosti, ktoré sa naučia, ako z údajov vyťažiť skryté vzorce a vytvoriť nové poznatky, si vytvoria udržateľnú konkurenčnú výhodu

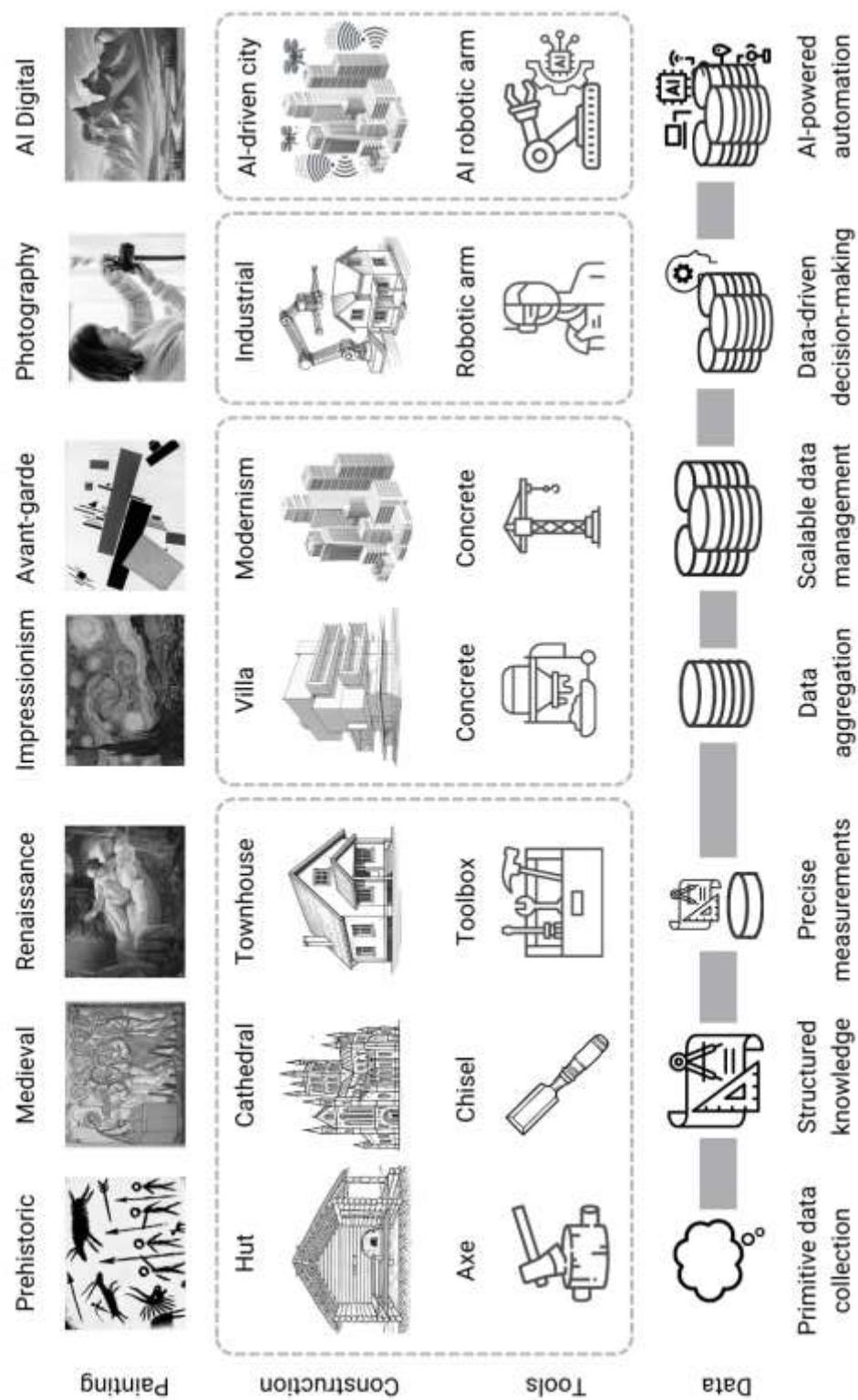
Hľadanie skrytých vzorcov a zmyslu údajov nie je len práca s číslami, ale tvorivý proces, ktorý si vyžaduje abstraktné myslenie a schopnosť vidieť za rôznorodými prvkami celý obraz. V tomto zmysle možno vývoj práce s údajmi prirovnáť k vývoju výtvarného umenia (obrázok 10.1-2).

Vývoj stavebníctva sa nápadne podobá vývoju výtvarného umenia. V oboch prípadoch ľudstvo pokročilo od primitívnych metód k sofistikovaným technológiám vizualizácie a analýzy. V praveku ľudia používali na riešenie každodenných úloh jaskynné kresby a primitívne nástroje. Počas stredoveku a renesancie sa úroveň sofistikovanosti architektúry a umenia dramaticky zvýšila. Do začiatku stredoveku sa stavebné nástroje vyvinuli z jednoduchej sekery na rozsiahle súpravy nástrojov, ktoré symbolizovali rast technických znalostí.

Obdobie realizmu bolo prvou revolúciou vo výtvarnom umení: umelci sa naučili reprodukovať aj tie najmenšie detaily a dosiahnuť maximálnu vieročnosť. V stavebníctve boli náprotivkom tohto obdobia presné inžinierske techniky, podrobne výkresy a prísne regulované výpočty, ktoré sa stali základom projektovej praxe na celé stáročia.

Neskôr impresionizmus zmenil samotné vnímanie umeleckej reality: namiesto doslovného vykreslovania formy začali umelci zachytávať náladu, svetlo a dynamiku, pričom sa snažili vyjadriť skôr celkový dojem než absolútну presnosť. Podobne aj strojové učenie v oblasti analýzy budov sa posúva od rigidných logických modelov k rozpoznávaniu vzorov a pravdepodobnostných modelov, ktoré umožňujú "vidieť" skryté závislosti v údajoch, ktoré sú pri klasickej analýze nedostupné. Tento prístup je ozvenou myšlienok minimalizmu a funkčnosti Bauhausu, kde je význam (funkcia) dôležitejší ako forma. Bauhaus sa snažil odstrániť nadbytočné, vzdať sa ornamentov v záujme jasnosti, utilitarizmu a masovej príťažlivosti. Veci museli byť zrozumiteľné a užitočné, bez zbytočnosti - estetika sa zrodila z logiky dizajnu a účelu.

S príchodom fotografie koncom 19. storočia získalo umenie nový nástroj na zachytenie reality s nevídanicou presnosťou a prevrátilo postoj k výtvarnému umeniu naruby. Podobne aj v stavebníctve priemyselná revolúcia v 21. storočí vedie k využívaniu robotických technológií, laserov, internetu vecí, RFID a konceptov, ako je Connected Construction, kde sa zber jednotlivých parametrov vyvinul do škálovateľného inteligentného zachytenia celej reality staveniska.



Obr. 10.1-2 Éra výtvarného vývoja je v súlade s vývojom prístupov k práci s údajmi v stavebníctve.

Tak ako dnes výtvarné umenie zažíva nový pohľad na svet s príchodom nástrojov umelej inteligencie a LLM, stavebníctvo zažíva ďalší skok: inteligentné systémy riadené umelou inteligenciou (AI), LLM chaty umožňujú predvídať, optimalizovať a generovať riešenia s minimálnym zásahom človeka.

Úloha údajov pri navrhovaní a riadení sa radikálne zmenila. Kým predtým sa poznatky prenášali ústne a mali empirický charakter - podobne ako sa realita až do 19. storočia zachytávala ručne maľovanými obrazmi - dnes sa pozornosť sústredí na kompletné digitálne zachytenie "obrazu" stavby. Pomocou algoritmov strojového učenia sa tento digitálny obraz transformuje na impresionistickú reprezentáciu stavebnej reality - nie na presnú kópiu, ale na zovšeobecnené, pravdepodobnostné chápanie procesov.

Rýchlo sa blížime do éry, v ktorej budú procesy navrhovania, výstavby a prevádzky budov nielen rozšírené, ale do veľkej miery riadené systémami umelej inteligencie. Podobne ako sa moderné digitálne umenie vytvára bez štetca - pomocou textových podnetov a generatívnych modelov - budú architektonické a inžinierske riešenia budúcnosti formované klúčovými požiadavkami a parametrami, ktoré nastaví používateľ'.

V 21. storočí sa prístup k údajom, ich interpretácia a kvalita analýzy stávajú nevyhnutnými pre úspech projektu. A hodnotu údajov neurčuje ich objem, ale schopnosť odborníkov ich analyzovať, overiť a premeniť na činy.

### **Prístup založený na údajoch v stavebníctve: nová úroveň infraštruktúry**

V histórii ľudstva priniesol každý takýto technologický skok zásadné zmeny v hospodárstve a spoločnosti. Dnes sme svedkami novej vlny transformácie, ktorá je svojím rozsahom porovnateľná s priemyselnou revolúciou v 19. storočí. Avšak kým pred sto rokmi boli hlavným motorom zmien mechanické sily a energetické technológie, teraz sú to dátá a umelá inteligencia.

Agenti strojového učenia, LLM a AI menia samotnú podstatu aplikácií, čím sa tradičné softvérové zásobníky (o ktorých sa hovorí v druhej časti knihy) stávajú zbytočnými (obr. 2.2-3). Všetka dátová logika je sústredená v agentoch AI, a nie v pevne zakódovaných obchodných pravidlach (obr. 2.2-4).

V dobe dátovej sa tradičný pohľad na aplikácie zásadne mení. Prechádzame na model, v ktorom objemné, modulárne podnikové systémy nevyhnutne ustúpia otvoreným, ľahkým, prispôsobeným riešeniam.

*V budúcnosti zostane len základná dátová štruktúra a všetka interakcia s ňou bude prebiehať prostredníctvom agentov pracujúcich priamo s databázou. Skutočne verím, že celý aplikačný zásobník zmizne, pretože ho umelá inteligencia pri priamej interakcii s podkladovou databázou jednoducho nebude potrebovať. Celú svoju kariéru som strávil prácou v SaaS - budoval som spoločnosti, pracoval som v nich, a ak mám byť úprimný, pravdepodobne by som teraz nespustil nový podnik SaaS. A pravdepodobne by som teraz neinvestoval do spoločností SaaS. Situácia je príliš neistá. Tým nechcem povedať, že v budúcnosti nebudú existovať softvérové spoločnosti, len budú vyzeráť úplne inak. Budúce systémy budú databázy s obchodnou logikou prenesenou do agentov [AI]. Títo agenti budú pracovať s viacerými dátovými úložiskami súčasne, nebudú sa obmedzovať na jednu databázu. Všetka logika sa presunie do vrstvy AI [46].*

- Matthew Berman, generálny riaditeľ spoločnosti Forward Future

Hlavným rozdielom novej paradigmy je minimalizácia technologického balastu. Namiesto monumentálnych zložitých a uzavretých softvérových systémov dostaneme flexibilné, otvorené a rýchlo prispôsobiteľné moduly, ktoré doslova "žijú" vo vnútri toku údajov (obr. 7.4-1 - Apache Airflow, NiFi). Architektúra budúceho riadenia procesov predpokladá používanie mikroaplikácií - kompaktných, účelovo vytvorených nástrojov, ktoré sa zásadne líšia od masívnych a uzavretých systémov ERP, PMIS, CDE, CAFM. Noví agenti budú čo najviac adaptívni, integrovaní a prispôsobení konkrétnemu podniku (napr. Low-Code/No-Code Obr. 7.4-6).

*Všetka obchodná logika bude smerovať k týmto agentom [AI] a títo agenti budú vykonávať operácie CRUD [Create, Read, Update, and Delete] na viacerých úložiskách, čo znamená, že nebudú rozlišovať, ktorý backend sa používa. Budú aktualizovať viacero databáz a všetka logika skončí v takzvanej AI -úrovni. A keď sa všetka logika bude nachádzať na úrovni AI, ľudia začnú vymieňať backendy. Už teraz vidíme dosť vysoké percento víťazstiev na trhu v oblasti používania backendov a agentov Dynamics a budeme sa agresívne uberať týmto smerom a snažiť sa to všetko spojiť. Či už v oblasti služieb zákazníkom alebo v iných oblastiach, napríklad nielen v CRM, ale aj v našich finančných a prevádzkových riešeniach. Ľudia totiž chcú viac obchodných aplikácií riadených umelou inteligenciou, kde logickú vrstvu môže riadiť umelá inteligencia a agenti s umelou inteligenciou. [...]. Jednou z najzaujímavejších vecí je pre mňa Excel s Pythonom, čo je porovnatelné s GitHubom s Copilotom. To znamená, čo sme urobili: teraz, keď máte Excel, mali by ste ho jednoducho otvoriť, spustiť Copilot a začať sa s ním hrať. Už to nie je len porozumenie dostupným číslam - sám vytvorí plán. Tak ako pracovná plocha GitHub Copilot vytvorí plán a potom ho vykoná, je to ako dátový analytik, ktorý používa Excel ako nástroj na vizualizáciu riadkov a stĺpcov na analýzu. Copilot teda používa Excel ako nástroj so všetkými jeho možnosťami, pretože dokáže generovať údaje a má interpreter jazyka Python.*

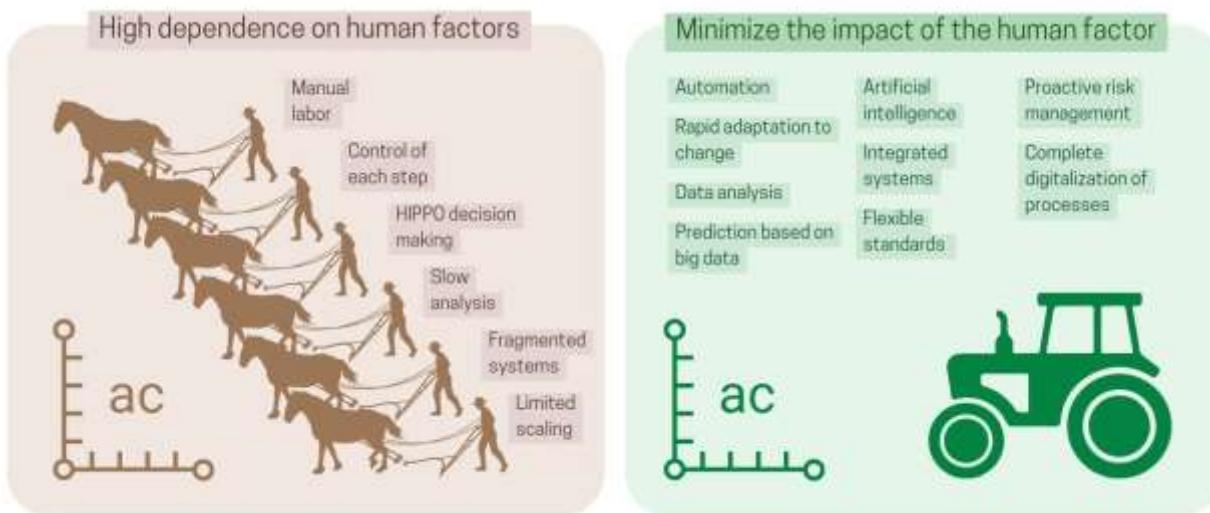
- Satya Nadella, generálny riaditeľ, Microsoft, rozhovor s kanálom BG2 december 2024. [28]

Premena, ktorej sme svedkami v logike kancelárskych aplikácií - prechod od modulárnych, uzavretých systémov k agentom umelej inteligencie pracujúcim priamo s otvorenými údajmi - je len súčasťou

ovel'a väčšieho procesu. Nejde len o zmenu rozhraní alebo softvérovej architektúry: zmeny ovplyvnia základné princípy organizácie práce, rozhodovania a riadenia podniku. V stavebníctve to povedie k logike založenej na údajoch, v ktorej sa údaje stanú ústredným prvkom procesov od projektovania až po riadenie zdrojov a monitorovanie výstavby.

## Digitálna kancelária novej generácie: ako AI mení pracovný priestor

Pred takmer sto rokmi už ľudstvo zažívalo podobnú technologickú revolúciu. Prechod od parných strojov k elektromotorom trval viac ako štyri desaťročia, ale v konečnom dôsledku katalyzoval nevídany rast produktivity - predovšetkým vďaka decentralizácii energetických kapacít a flexibilite nových riešení. Tento posun nielenže zmenil chod dejín a presunul väčšinu obyvateľstva z vidieka do miest, ale položil aj základy moderného hospodárstva. História technológií je cestou od fyzickej práce k automatizácii a inteligentným systémom. Tak ako traktor nahradil desiatky obrábačov pôdy, moderné digitálne technológie nahrádzajú tradičné kancelárske metódy riadenia výstavby (obrázok 10.1-3). Ešte na začiatku 20. storočia väčšina svetovej populácie obrábala pôdu ručne, až kým sa v 30. rokoch 20. storočia nezačala mechanizácia práce pomocou strojov a traktorov.



Obr. 10.1-3 Tak ako traktor nahradil na začiatku 20. storočia desiatky ľudí, strojové učenie nahradí v 21. storočí tradičné metódy riadenia podnikov a projektov.

Tak ako ľudstvo pred sto rokmi prešlo od obrábania jednotlivých pozemkov pomocou primitívnych nástrojov k veľkoplošnému poľnohospodárstvu s využitím strojov, dnes prechádzame od spracovania rôznorodých "síl" informácií k práci s dátovými poliami pomocou výkonných "traktorov" - ETL-pipeline a algoritmov umelej inteligencie.

Sme na prahu podobného skoku - ale v digitálnej rovine: od tradičného manuálneho riadenia podniku k modelom založeným na údajoch.

Cesta k plnohodnotnej architektúre založenej na údajoch si bude vyžadovať čas, investície a organizačné úsilie. Táto cesta však otvára cestu nielen k postupnému zlepšovaniu, ale ku kvalitatívному skoku smerom k vyššej efektívnosti, transparentnosti a riaditeľnosti stavebných

procesov. To všetko je podmienené systematickým zavádzaním digitálnych nástrojov a opustením zastaraných obchodných postupov.

Parametrizácia úloh, ETL, LLM, komponenty IoT, RFID, tokenizácia, big **data** a strojové učenie zmenia tradičné stavebníctvo na **stavebníctvo riadené údajmi**, kde sa každý detail projektu a stavebného podniku bude riadiť a optimalizovať pomocou údajov.

Analýza informácií si predtým vyžadovala tisíce človekohodín. Teraz tieto úlohy vykonávajú algoritmy a LLM, ktoré pomocou podnetov menia rôznorodé súbory údajov na strategické zdroje. Vo svete technológií sa deje to isté, čo sa stalo v poľnohospodárstve: prechádzame od motyk k automatizovanému agrotechnickému podnikaniu. Tak aj kancelárska práca v stavebníctve - od súborov Excel a manuálneho sumarizovania - prechádza na inteligentný systém, v ktorom sa údaje zhromažďujú, čistia, štruktúrujú a menia na poznatky.

Už dnes by mali spoločnosti začať "pestovať" informačné polia prostredníctvom kvalitného zberu údajov a štruktúrovania informácií, "hnojiť" ich čistiacimi a normalizačnými nástrojmi a následne ich "zbierať" v podobe prediktívnej analýzy a automatizovaných riešení. Ak moderný farmár so strojom dokáže nahrať sto obrábačov pôdy, inteligentné algoritmy budú schopné zbaviť zamestnancov rutiny a previesť ich do úlohy strategických manažérov informačných tokov.

Je však dôležité si uvedomiť, že vytvorenie skutočne dátovo orientovanej -organizácie nie je rýchly proces. Je to dlhodobé strategické smerovanie, podobné vytváraniu nového miesta na pestovanie nového lesa (obr. 1.2-5) systémov, kde každý "strom" v tomto ekosystéme je samostatný proces, kompetencia alebo nástroj, ktorý potrebuje čas na rast a rozvoj. Podobne ako v prípade skutočného lesa, úspech závisí nielen od kvality sadbového materiálu (technológie), ale aj od pôdy (firemnnej kultúry), klímy (podnikateľského prostredia) a starostlivosti (systémového prístupu).

Spoločnosti sa už nebudú môcť spoliehať len na uzavreté, hotové riešenia. Na rozdiel od predchádzajúcich etáp technologického vývoja je nepravdepodobné, že by súčasný prechod - k otvorenému prístupu k údajom, umelej inteligencii a otvorenému softvéru - podporovali veľkí dodávatelia, pretože priamo ohrozujú ich zavedené obchodné modely a hlavné zdroje príjmov.

Ako ukázala štúdia Harvard Business School [40], o ktorej sa už hovorilo v kapitole o štvrtej a piatej technologickej revolúcii, náklady na vytvorenie najpoužívanejších Open Source riešení od začiatku pre všetky spoločnosti by boli približne 4,15 miliardy dolárov. Ak si však predstavíme, že by si každá spoločnosť vyvinula vlastné alternatívy bez prístupu k existujúcim nástrojom Open Source, čo sa deje v posledných desaťročiach, celkové náklady podnikov by mohli dosiahnuť kolosalných 8,8 bilióna dolárov - to je cena iracionálneho dopytu, na ktorú možno oceniť trh so softvérom.

Technologický pokrok nevyhnutne povedie k prehodnoteniu zavedených obchodných modelov. Kým predtým mohli spoločnosti zarábať na zložitých, neprehľadných procesoch a uzavretých údajoch, s rozvojom umelej inteligencie a analytyky je tento prístup čoraz menej životoschopný.

V dôsledku demokratizácie prístupu k údajom a nástrojom sa tradičný trh predaja softvéru môže výrazne zmeniť. Zároveň však porastie nový trh - trh digitálnych odborných znalostí, prispôsobenia, integrácie a návrhu riešení. V tomto prípade nebude hodnota pochádzať z predaja licencií, ale zo schopnosti vytvárať flexibilné, otvorené a prispôsobivé digitálne procesy. Tak ako elektrifikácia a nástup traktorov dali vzniknúť novým priemyselným odvetviám, tak aj aplikácia veľkých dát, umelej inteligencie a LLM otvorí úplne nové obzory podnikania v stavebnictve, ktoré si budú vyžadovať nielen technologické investície, ale aj hlbokú transformáciu myslenia, procesov a organizačných štruktúr. A tie spoločnosti a odborníci, ktorí si to uvedomia a začnú konať už dnes, budú lídrami zajtrajska.

Vo svete, kde sa otvorené údaje stávajú hlavným prínosom, bude dostupnosť informácií meniť pravidlá hry. Investori, klienti a regulačné orgány budú čoraz viac požadovať transparentnosť a algoritmy strojového učenia budú schopné automaticky identifikovať nezrovnalosti v odhadoch, časových harmonogramoch a nákladoch. To vytvára podmienky pre novú etapu digitálnej transformácie, ktorá nás postupne vedie k "uberizácii" stavebníctva.

## Otvorené údaje a uberizácia sú hrozbou pre existujúce stavebné podniky

Stavebníctvo sa stáva procesom riadenia informácií. Čím presnejšie, kvalitnejšie a úplnejšie sú údaje, tým efektívnejšie je navrhovanie, výpočty, odhady nákladov, výstavba a prevádzka budov. V budúcnosti nebude kľúčovým zdrojom žeriav, betón a výstuž, ale schopnosť zhromažďovať, analyzovať a využívať informácie.

Klienti stavebných spoločností - investori a klienti financujúci výstavbu - budú v budúcnosti nevyhnutne využívať hodnotu otvorených údajov a analytiky historických údajov. Otvoria sa tak možnosti automatizácie výpočtu časového harmonogramu projektu a nákladov bez toho, aby sa stavebné spoločnosti zaoberali otázkami kalkulácie nákladov, čo pomôže kontrolovať náklady a rýchlejšie identifikovať nadbytočné náklady.

Predstavte si stavenisko, na ktorom laserové skenery, kvadrokoptéry a fotogrametrické systémy zbierajú presné údaje o množstve použitého betónu v reálnom čase. Tieto informácie sa automaticky konvertujú do jednoduchých plochých modelov MESH s metadátami, čím sa obchádzajú ťažkopádne systémy CAD (BIM) bez závislosti od zložitých geometrických jadier, ERP alebo PMIS. Tieto údaje zozbierané zo staveniska sa centrálnie prenášajú do jedného štruktúrovaného úložiska, ktoré je k dispozícii klientovi na nezávislú analýzu, kde sa nahrávajú reálne ceny z rôznych stavebných obchodov a parametre od miery úverového financovania až po dynamicky sa meniace faktory, ako sú poveternostné podmienky, burzové kotácie stavebných materiálov, logistické tarify a štatistické sezónne výkyvy cien práce. Za takýchto podmienok sú okamžite zrejmé akékoľvek rozdiely medzi projektovanými a skutočnými objemami materiálu, čo znemožňuje manipuláciu s odhadmi vo fáze projektovania aj pri odovzdávaní projektu. Výsledkom je, že transparentnosť stavebného procesu sa nedosahuje prostredníctvom armády dozorcov a manažérov, ale prostredníctvom objektívnych

digitálnych údajov, ktoré minimalizujú ľudský faktor a možnosť špekulácií.

V budúcnosti budú tento druh kontroly údajov vykonávať skôr správcovia údajov na strane zákazníka (obr. 1.2-4 CQMS manager). Platí to najmä pre kalkulácie a odhady projektov: tam, kde doteraz existovalo celé oddelenie odhadcov, budú už zajtra existovať nástroje strojového učenia a prognostické nástroje, ktoré budú určovať cenové limity, do ktorých sa stavebné firmy zmestia.

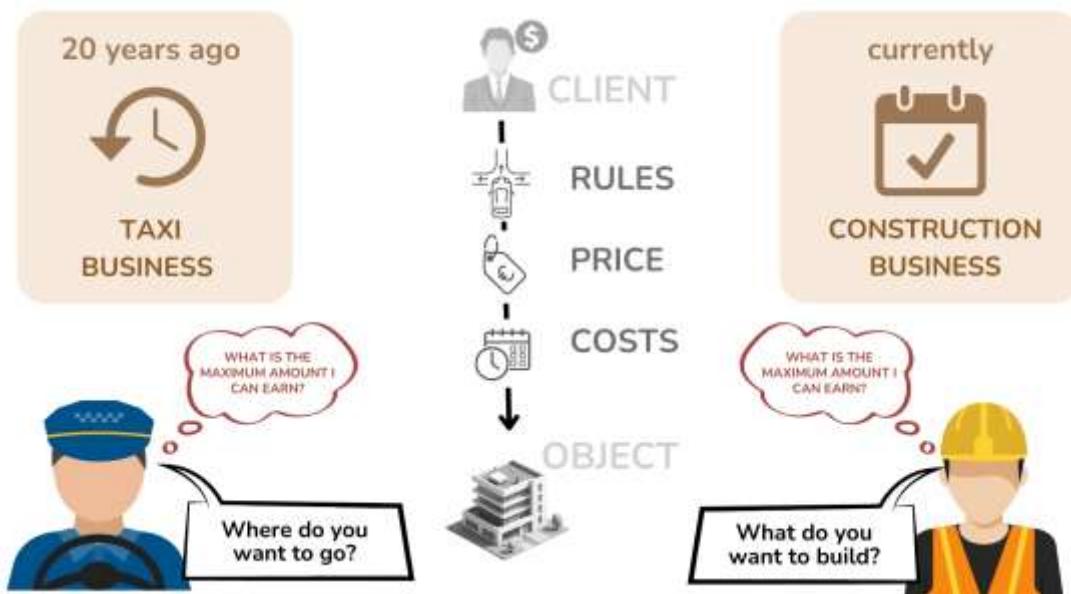
*Vzhľadom na roztrieštenosť [stavebného] priemyslu, kde väčšinu systémov a subsystémov dodávajú malé a stredné podniky, musí digitálna stratégia vychádzať od zákazníka. Zákazníci musia vytvoriť podmienky a mechanizmy na uvoľnenie digitálnych schopností dodávateľského reťazca [20].*

- Andrew Davis a Giuliano Denicol, Accenture "Vytváranie vyššej hodnoty prostredníctvom kapitálových projektov"

Takáto otvorenosť a transparentnosť údajov predstavuje hrozbu pre stavebné spoločnosti, ktoré sú zvyknuté zarábať na neprehľadných procesoch a mätúcich správach, kde sa špekulácie a skryté náklady môžu skrývať za zložitými a uzavretými formátmi a modulárnymi proprietárnymi dátovými platformami. Preto je nepravdepodobné, že by stavebné spoločnosti, podobne ako dodávatelia propagujúci riešenia s otvoreným zdrojovým kódom, mali záujem o úplnú implementáciu otvorených údajov do svojich obchodných procesov. Ak sú údaje dostupné a ľahko spracovateľné pre zákazníka, môžu sa kontrolovať automaticky, čím sa eliminuje možnosť nadhodnotenia objemov a manipulácie s odhadmi.

Podľa správy Svetového ekonomickeho fóra "Formovanie budúcnosti stavebníctva" (2016) [5] zostáva jednou z kľúčových výziev pre toto odvetvie pasívna úloha klienta. Napriek tomu sú to práve zákazníci, ktorí by mali prevziať väčšiu zodpovednosť za výsledok projektov - od včasného plánovania, cez výber udržateľných modelov interakcie až po monitorovanie výkonnosti. Bez aktívnej účasti vlastníkov projektov nie je možná systémová transformácia stavebníctva.

Strata kontroly nad objemom a výpočtom nákladov už za posledných 20 rokov zmenila iné odvetvia, pričom umožnila zákazníkom priamo, bez sprostredkovateľov, stymalizovať svoje ciele. Digitalizácia a transparentnosť údajov zmenili mnohé tradičné obchodné modely, napríklad taxikárov s príchodom Uberu (obrázok 10.1-4), hotelierov s príchodom Airbnb a maloobchodníkov a obchodov s príchodom Amazonu a banky s príchodom nebanskoviek a decentralizovaných fintech ekosystémov, kde priamy prístup k informáciám a automatizácia výpočtov času a nákladov výrazne znížili úlohu sprostredkovateľov.



Obr. 10.1-4 Stavebníctvo bude čeliť überizácii, akej museli pred 10 rokmi čeliť taxikári, hotelieri a maloobchodníci.

Proces demokratizácie prístupu k údajom a nástrojom na ich spracovanie je nevyhnutný a časom sa otvorené údaje o všetkých zložkách projektu stanú požiadavkou zákazníka a novým štandardom. Preto budú otázky zavedenia otvorených formátov a transparentných výpočtov presadzovať investori, zákazníci, banky a fondy súkromného kapitálu (private equity) - teda tí, ktorí sú koncovými používateľmi vybudovaných objektov a následne objekt prevádzkujú desiatky rokov.

Veľkí investori, klienti a banky už požadujú transparentnosť v stavebníctve. Podľa štúdie spoločnosti Accenture "Vytváranie väčšej hodnoty prostredníctvom kapitálových projektov" (2020) [20] sa transparentné a spoľahlivé údaje stávajú rozhodujúcim faktorom pre investičné rozhodnutia v stavebníctve. Ako zdôrazňujú odborníci, dôveryhodné a efektívne riadenie projektov nie je možné bez transparentnosti, najmä v čase krízy. Okrem toho vlastníci aktív a dodávateelia čoraz viac prechádzajú na zmluvy, ktoré motivujú k zdielaniu údajov a spoločnej analýze, čo odráža rastúce požiadavky investorov, bánsk a regulačných orgánov na zodpovednosť a transparentnosť.

Pohyb investora, zákazníka od nápadu k hotovej stavbe, sa v budúcnosti bude podobat cestovaniu na autopilota - bez vodiča v podobe stavebnej firmy, sľubuje, že sa stane nezávislým od špekulácií a neistoty.

Éra otvorených údajov a automatizácie nevyhnutne zmení stavebníctvo, podobne ako sa to už stalo v bankovníctve, obchode, poľnohospodárstve a logistike. V týchto odvetviach úloha sprostredkovateľov a tradičné spôsoby podnikania ustupujú automatizácii a robotizácii, pričom nezostáva priestor na neodôvodnené prirážky a špekulácie.

Údaje a procesy vo všetkých hospodárskych činnostiach človeka sa nelisia od tých, s ktorými sa musia

zaoberať odborníci v stavebníctve. Z dlhodobého hľadiska môžu stavebné spoločnosti, ktoré dnes dominujú na trhu tým, že určujú ceny a štandardy kvality služieb, stratiť svoju úlohu kľúčového sprostredkovateľa medzi zákazníkom a jeho stavebným projektom.

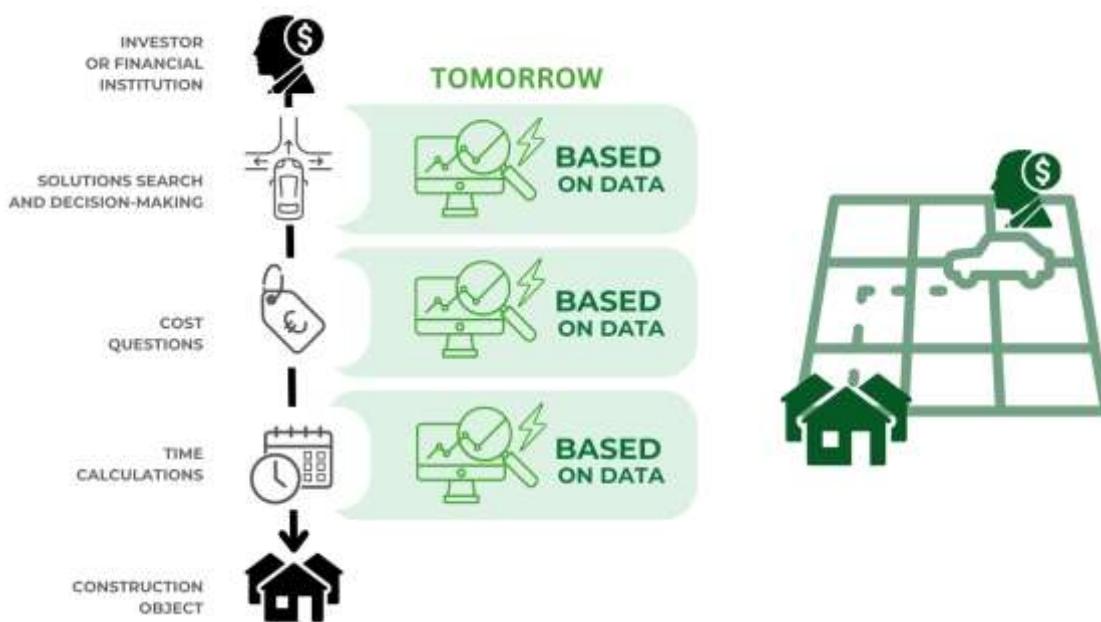
## **Nevyriešené problémy uberezácie ako posledná šanca využiť čas na transformáciu**

Vráťme sa však k realite stavebníctva. Zatiaľ čo v niektorých odvetviach hospodárstva sa objavujú samojazdiace autá, decentralizované finančné systémy a riešenia založené na umelej inteligencii, značná časť stavebných spoločností stále zostáva organizáciami založenými na papierovej forme, v ktorých sa kľúčové rozhodnutia prijímajú skôr na základe intuície a skúseností jednotlivých odborníkov.

V tejto paradigmе možno modernú stavebnú spoločnosť prirovnáť k 20 rokov starej taxislužbe, ktorá kontroluje zdroje, trasy a časy dodávok a je zodpovedná za načasovanie a náklady "cesty" - od myšlienky projektu (logistiky a procesu montáže) až po dodanie projektu. Tak ako GPS (v stavebníctve IoT, RFID) a algoritmy strojového učenia pri výpočtoch času/nákladov kedysi zmenili oblasť dopravy, údaje, algoritmy a umelá inteligencia -agenti majú potenciál zmeniť riadenie stavebníctva - od intuitívneho hodnotenia k prediktívnym, riadeným modelom. Za posledných 20 rokov mnohé odvetvia - financie, polnohospodárstvo, maloobchod a logistika - postupne eliminovali možnosť špekulovať vďaka neprehľadnosti údajov. Ceny, náklady na dodávky alebo finančné transakcie sa vypočítavajú automaticky a štatisticky podložene - v priebehu niekoľkých sekúnd na digitálnych platformách.

Pri pohľade do budúcnosti si stavebné spoločnosti musia uvedomiť, že demokratizácia prístupu k údajom a nástrojom na ich analýzu naruší tradičný prístup k odhadovaniu nákladov a časového harmonogramu projektu a odstráni možnosť špekulovať na základe neprehľadných údajov o objeme a cene.

Podobne ako jazda po regulovanej ceste bez zásahu vodiča, aj stavebné procesy budúcnosti budú čoraz viac pripomínať systém "Uber" - s automatizovaným odhadom času a nákladov, transparentným smerovaním úloh a minimálnou závislosťou od ľudského faktora. Tým sa zmení samotná povaha "cesty" od nápadu po realizáciu - stane sa predvídateľnejšou, lepšie riadenou a riadenou údajmi.



Obr. 10.1-5 Náklady a "cestovný" čas počas výstavby sa určia pomocou strojového učenia a štatistických nástrojov.

S postupným zavádzaním nových predpisov a požiadaviek takmer v každej krajine sveta, ktoré ukladajú povinnosť odovzdávať CAD- (BIM)-modely klientom alebo bankám financujúcim stavebné projekty, majú klient a zákazník možnosť nezávisle zabezpečiť transparentnosť kalkulácií nákladov a rozsahu. Je to dôležité najmä pre veľkých zákazníkov a investorov, ktorí majú dostatočné kompetencie a nástroje na promptnú analýzu objemov a monitorovanie trhových cien. Pre spoločnosti, ktoré realizujú veľké štandardné projekty - obchody, administratívne budovy, obytné komplexy - sa takéto postupy stávajú štandardom.

Kedže informačný obsah modelov sa stáva úplnejším a štandardizovanejším, možnosť manipulácie a špekulácie prakticky mizne. Digitálna transformácia postupne mení pravidlá hry v stavebníctve a spoločnosti, ktoré sa týmto zmenám neprispôsobia, môžu čeliť vážnym problémom.

Zvýšená konkurencia, technologické narušenie a znižujúce sa marže majú potenciál ovplyvniť udržateľnosť podnikania. Vzhľadom na obmedzenú likviditu sa čoraz viac subjektov v odvetví obracia na automatizáciu, analytiku a dátové technológie ako na spôsob, ako zvýšiť efektívnosť a transparentnosť procesov. Tieto nástroje sa stávajú dôležitým zdrojom na udržanie konkurenčieschopnosti v meniacom sa ekonomickom prostredí.

Možno by sme nemali čakať, kým si vonkajšie okolnosti vynútia naliehavé opatrenia - oveľa efektívnejšie je začať sa pripravovať už dnes posilňovaním digitálnych kompetencií, zavádzaním moderných riešení a budovaním kultúry orientovanej na údaje.

Jednou z posledných kľúčových technologických prekážok rozsiahlej digitálnej transformácie stavebníctva, ktorá sa v najbližších rokoch dotkne každej spoločnosti, je problém automatickej klasifikácie prvkov stavebných projektov.

Bez spoločnej, presnej a škálovateľnej klasifikácie nie je možné vytvoriť základ pre plnohodnotnú analýzu, automatizáciu procesov a riadenie životného cyklu pomocou umej inteligencie a prediktívnych modelov. Pokial klasifikácia objektov stále závisí od manuálnej interpretácie skúsenými odborníkmi - majstrami, projektantmi, odhadcami -, stavebníctvo má stále príležitosť. Tento čas možno využiť na prípravu na nevyhnutné zmeny: rastúce požiadavky na transparentnosť, demokratizáciu nástrojov a údajov a nástup automatických klasifikačných systémov, ktoré radikálne zmenia pravidlá hry.

Úloha automatickej klasifikácie prvkov stavebného sveta je svojou zložitosťou porovnateľná s rozpoznávaním objektov v bezpilotných jazdných systémoch, čo je jedna z hlavných výziev. Predstavme si bezpilotné vozidlo pohybujúce sa z bodu A do bodu B (obr. 10.1-5). Súčasné systémy automatickej jazdy sa potýkajú s problémom klasifikácie objektov, ktoré sú rozpoznávané lidmi a kamerami. Nestačí, aby auto jednoducho "videlo" prekážku alebo orientačný bod, ale musí byť schopné rozpoznať, či ide o chodca, dopravnú značku alebo odpadkový kôš.

Podobnej zásadnej výzve čelí celé stavebníctvo. Prvky projektu - ako sú okná, dvere alebo stípy - môžu byť zaznamenané v dokumentácii, znázornené v modeloch CAD, odfotografované na stavbe alebo rozpoznané v mračnách bodov z laserového skenovania. Ich vizuálne alebo hrubé geometrické rozpoznanie však nestačí na vytvorenie skutočne automatizovaného systému riadenia projektov. Je potrebné zabezpečiť, aby bol každý prvak presne a dôsledne zaradený do typu, ktorý bude jednoznačne identifikovateľný vo všetkých následných procesoch - od odhadov a špecifikácií až po logistiku, kontrolu zásob a predovšetkým prevádzku (obr. 4.2-6).

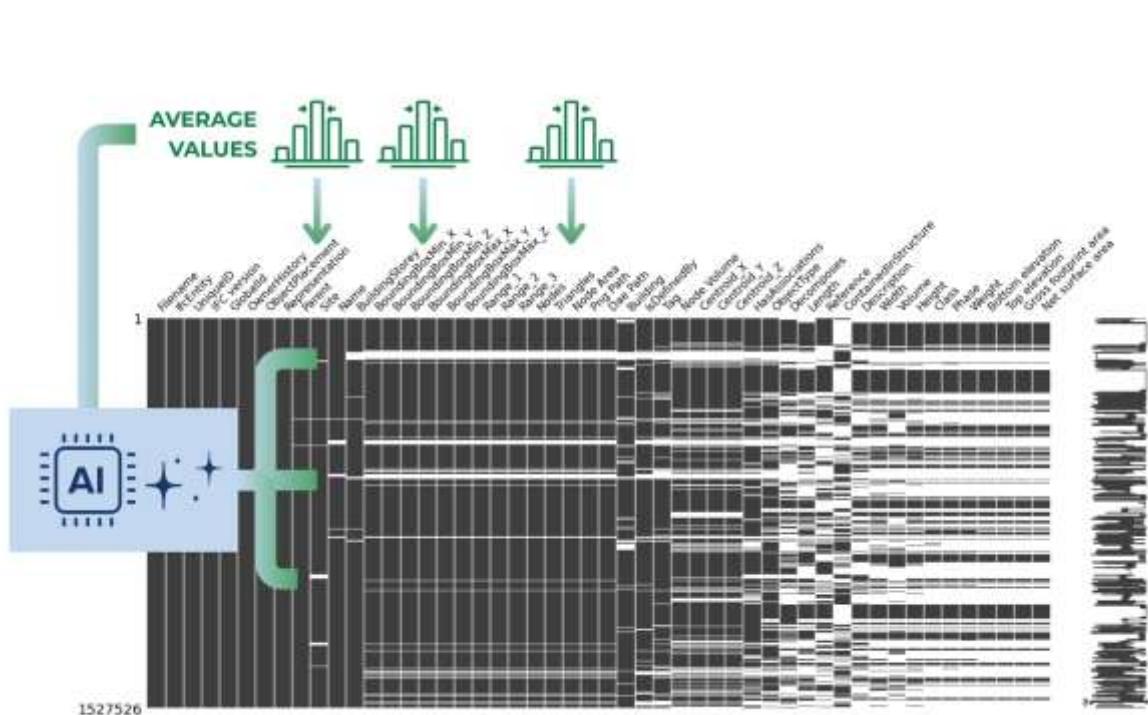
Práve v tejto fáze - prechod od rozpoznania k zmysluplnnej klasifikácii - vzniká jedna z kľúčových prekážok. Aj keď sú digitálne systémy technicky schopné rozlišovať a identifikovať objekty v modeloch a na stavenisku, hlavný problém spočíva v správnej a kontextovo stabilnej definícii typu prvku pre rôzne softvérové prostredia.. Napríklad dvere môže projektant v modeli CAD označiť ako prvak kategórie "dvere", ale pri prenose do systému ERP alebo PMIS môžu byť nesprávne označené, a to buď v dôsledku chyby projektanta, alebo v dôsledku nekonzistencia medzi systémami. Okrem toho prvak často stráca niektoré dôležité atribúty alebo počas exportu a importu údajov zmizne z účtovníctva systému. To vedie k medzere v toku údajov a narúša zásadu komplexnej digitalizácie stavebných procesov. Vzniká tak kritická medzera medzi "viditeľným" a "zrozumiteľným" sémantickým významom, ktorá narúša integritu údajov a výrazne komplikuje automatizáciu procesov počas celého životného cyklu stavebného projektu.

Vyriešenie problému univerzálnej klasifikácie stavebných prvkov pomocou technológií veľkých dát a strojového učenia (obr. 10.1-6) bude katalyzátorom transformácie celého odvetvia - a možno aj nečakaným objavom pre mnohé stavebné spoločnosti. Jednotný, učenlivý klasifikačný systém bude základom pre škálovateľnú analytiku, digitálne riadenie a prijatie umej inteligencie do každodennej praxe stavebných organizácií.

NVIDIA a ďalší technologickí lídri už poskytujú riešenia v iných odvetviach, ktoré dokážu automaticky kategorizovať a štruktúrovať obrovské množstvo textových a vizuálnych informácií.

Napríklad model NeMo Curator spoločnosti NVIDIA [161] sa špecializuje na automatickú klasifikáciu a kategorizáciu údajov do vopred definovaných kategórií, čo zohráva kľúčovú úlohu pri optimalizácii potrubí na spracovanie informácií na jemné ladenie a predtrénovanie generatívnych modelov umelej inteligencie. Platforma Cosmos sa trénuje na reálnych videách a 3D scénach [162], čím poskytuje základ pre autonómne systémy a digitálne dvojčaťa, ktoré sa už budujú v ekosystéme NVIDIA. NVIDIA Omniverse, ktorá sa do roku 2025 stala popredným nástrojom na prácu s formátom USD, univerzálnym popisom scény, ktorý by mohol časom nahradíť formát IFC v procesoch prenosu informácií o návrhu. Spolu s Isaac Sim, simulátorom robotických procesov [163], predstavujú riešenia ako NeMo Curator, Cosmos a Omniverse novú úroveň automatizácie: od čistenia a filtrovania údajov až po generovanie tréningových súborov, modelovanie vlastností objektov a tréning robotov na stavbe. A všetky tieto nástroje sú bezplatné a s otvoreným zdrojovým kódom, čo výrazne znižuje prekážky ich prijatia v inžinierskej a stavebnej praxi.

Automatická klasifikácia údajov na úrovni štruktúrovaných tabuľiek nie je taká zložitá, ako sa môže na prvý pohľad zdať. Ako sme ukázali v predchádzajúcej kapitole (obr. 9.1-10), ak existujú nahromadené historické údaje, je možné doplniť chýbajúce alebo nesprávne hodnoty tried na základe podobných parametrov iných prvkov. Ak už boli prvky s podobnými vlastnosťami správne klasifikované vo viacerých ukončených projektoch, systém môže s vysokou pravdepodobnosťou navrhnuť vhodnú hodnotu pre nový alebo neúplný prvek (obr. 10.1-6). Takáto logika založená na spriemerovaných hodnotách a analýze kontextu môže byť obzvlášť účinná pri hromadnom spracovaní tabuľkových údajov pochádzajúcich z odhadov, špecifikácií alebo modelov CAD.

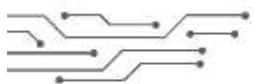


Obr. 10.1-6 Strojové učenie vám môže pomôcť automaticky nájsť priemerné hodnoty pre nevyplnené (biele polia) parametre tabuľky na základe minulých projektov.

Na pozadí takého rýchleho pokroku v strojovom učení je jasné: v roku 2025 je naivné veriť, že problém automatickej klasifikácie stavebných prvkov zostane dlho nevyriešený. Áno, súčasné algoritmy ešte nedosiahli úplnú vyspelosť, najmä pri neúplných alebo heterogénnych údajoch, ale okno príležitostí na prispôsobenie sa rýchlo uzatvára.

Spoločnosti, ktoré už investujú do zberu, čistenia a organizovania svojich údajov, ako aj do nástrojov na automatizáciu ETL, budú vo výhode. Ostatným hrozí, že budú zaostávať - rovnako ako kedysi spoločnosti nezvládli výzvy digitálnej transformácie v dopravnom a finančnom priemysle.

Tí, ktorí sa naďalej spoliehajú na manuálnu správu údajov a tradičné metódy odhadu nákladov a času, riskujú, že sa začiatkom roka 2020 ocitnú v pozícii vozových parkov taxislužieb z roku 2000, ktoré sa nedokážu prispôsobiť ére mobilných aplikácií a automatizovaných výpočtov tras.



## KAPITOLA 10.2.

### PRAKTICKÁ PRÍRUČKA NA ZAVEDENIE PRÍSTUPU ZALOŽENÉHO NA ÚDAJOCH

#### Od teórie k praxi: plán digitálnej transformácie v stavebníctve

Stavebníctvo postupne vstupuje do novej fázy vývoja, v ktorej sa známe procesy čoraz viac dopĺňajú - a niekedy aj nahradzajú - digitálnymi platformami a transparentnými interakčnými modelmi. To predstavuje pre spoločnosti nielen výzvy, ale aj významné príležitosti. Tie organizácie, ktoré už dnes budujú dlhodobú digitálnu stratégiju, si budú môcť nielen udržať svoju pozíciu na trhu, ale ju aj rozšíriť tým, že zákazníkom ponúknu moderné prístupy a spoľahlivé, technologicky podporené riešenia.

Je dôležité uvedomiť si, že znalosť pojmov a technológií je len východiskom. Manažéri a odborníci stoja pred praktickou otázkou: kde začať s implementáciou a ako premeniť teoretické myšlienky na skutočnú hodnotu. Okrem toho sa čoraz častejšie vynára otázka: na čom bude podnikanie založené, ak tradičné metódy kalkulácie nákladov a časového rozvrhu môže zákazník kedykoľvek prehodnotiť.

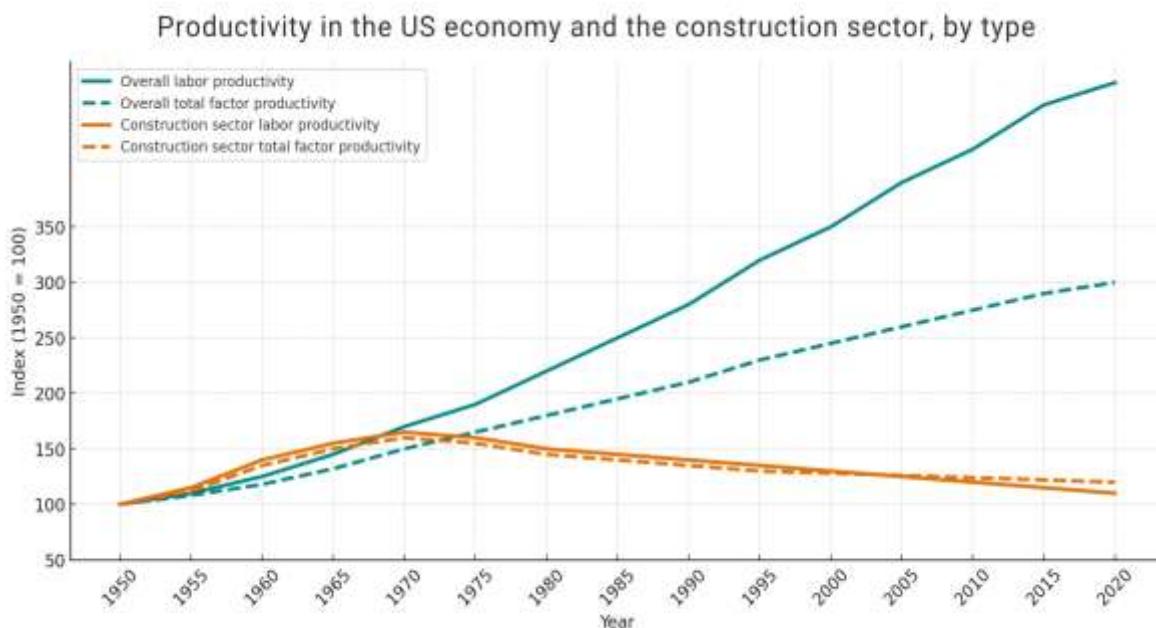
Odpoveď pravdepodobne nespočíva ani tak v technológii, ale vo formovaní novej profesionálnej kultúry, v ktorej sa práca s údajmi vníma ako neoddeliteľná súčasť každodennej praxe. Práve nedostatočná pozornosť venovaná digitálnym technológiám a inováciám naočkovala stavebníctvo do vážneho zaostávania, ktoré sa prejavuje v posledných desaťročiach [43].

Podľa spoločnosti McKinsey predstavujú výdavky na výskum a vývoj v stavebníctve menej ako 1% príjmov, zatiaľ čo v automobilovom a leteckom priemysle dosahujú 3,5 až 4,5%. Podobne aj náklady na IT v stavebníctve zostávajú na úrovni menej ako 1% celkových príjmov [107].

V dôsledku toho klesá nielen úroveň automatizácie, ale aj produktivita práce v stavebníctve a do roku 2020 bude stavebný robotník vyrábať menej ako pred polstoročím (obr. 10.2-1)

Takéto problémy s produktivitou v stavebníctve sú spoločné pre väčšinu rozvinutých aj rozvojových krajín (produktivita v stavebníctve klesla v 16 z 29 krajín OECD (obrázok 2.2-1)) a poukazujú nielen na nedostatok technológií, ale aj na potrebu systémových zmien v samotných prístupoch k riadeniu, vzdelávaniu a inováciám.

Úspech digitálnej transformácie nezávisí ani tak od počtu a dostupnosti nástrojov, ale od schopnosti organizácií prehodnotiť svoje procesy a vytvoriť kultúru otvorenú zmenám. Klúčová nie je samotná technológia, ale ľudia a procesy, ktoré zabezpečujú jej efektívne využívanie, podporujú neustále učenie a podnecujú prijímanie nových myšlienok.



Obr. 10.2-1 Paradox produktivity práce a celkovej produktivity zdrojov v ekonomike USA a v sektore stavebníctva (1950-2020) (na základe [43]).

V prvých častiach knihy bol model podnikateľského prostredia prirovnaný k lesnému ekosystému (obrázok 2.1-2, obrázok 1.2-4, obrázok 1.3-2). V zdravom lese zohrávajú periodické požiare pri všetkej svojej ničivej sile kľúčovú úlohu pri dlhodobej obnove. Čistia pôdu od starého porastu, vracajú do nej uložené živiny a vytvárajú priestor pre nový život. Niektoré rastlinné druhy sa dokonca vyvinuli tak, že ich semená sa otvárajú len vtedy, keď sú vystavené vysokým teplotám požiaru - prirodzený mechanizmus, ktorý poskytuje ideálny čas na klíčenie.

Podobne aj v podnikaní môžu krízy pôsobiť ako "riadené vyhorenie", ktoré podporuje vznik nových prístupov a spoločností, ktoré nie sú viazané na zastarané systémy. Takéto obdobia nútia opustiť neefektívne postupy, čím sa uvoľnia zdroje pre inovácie. Tak ako les po požiari začína pionierskymi rastlinami, tak aj podnik po kríze formuje nové, flexibilné procesy, ktoré sa stávajú základom vyspelého informačného prostredia.

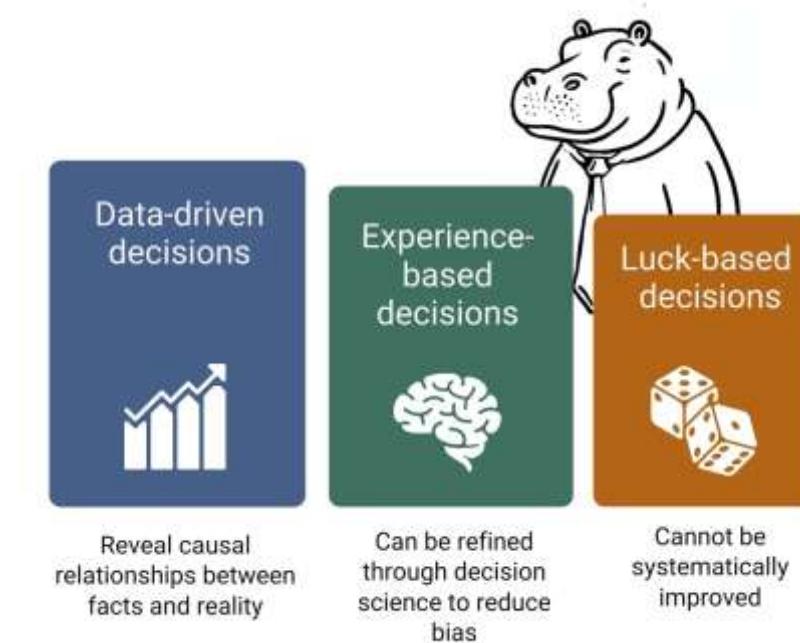
Spoločnosti, ktoré dokážu správne interpretovať tieto "signálne ohne" a transformovať ich rušivú energiu na konštruktívnu zmenu, dosiahnu novú úroveň výkonnosti - s transparentnejšími, adaptívnejšími dátovými procesmi, ktoré zvyšujú prirodzenú schopnosť organizácie obnovovať sa a rásť.

O rastúcom vplyve umelej inteligencie a strojového učenia na podnikateľské prostredie už nie je pochyb. Nejde len o dočasný trend, ale o strategickú nevyhnutnosť. Spoločnosti, ktoré ignorujú AI, riskujú stratu konkurenčieschopnosti na trhu, ktorý čoraz viac podporuje inovácie a flexibilitu.

Budúcnosť patrí tým, ktorí umelú inteligenciu nevnímajú len ako nástroj, ale ako príležitosť prehodnotiť každý aspekt svojho podnikania - od optimalizácie procesov až po manažérské rozhodnutia.

## Položenie digitálneho základu: 1-5 krokov k digitálnej zrelosti

V tejto kapitole sa pozrieme na plán digitálnej transformácie a identifikujeme kľúčové kroky potrebné na zavedenie prístupu založeného na údajoch, ktorý môže pomôcť transformovať podnikovú kultúru aj informačný ekosystém spoločnosti.



Obrázok 10.2-2 Riadená aktualizácia a výber stratégie: prípad, skúsenosť alebo údaje.

Podľa štúdie spoločnosti McKinsey "Prečo digitálne stratégie zlyhávajú" (2018) existuje najmenej päť dôvodov [164], prečo spoločnosti nedosahujú ciele digitálnej transformácie

- **Nejasné definície:** Vedúci pracovníci a manažéri majú rôzne chápanie toho, čo znamená pojem "digitálny", čo vedie k nedorozumeniam a nezrovnalostiam.
- **Nepochopenie digitálnej ekonomiky:** mnohé spoločnosti podceňujú rozsah zmien, ktoré digitalizácia prináša do obchodných modelov a dynamiky odvetvia (obrázok 10.1-6).
- **Ignorovanie ekosystémov:** spoločnosti sa zameriavajú na jednotlivé technologické riešenia (dátové silá) a prehliadajú potrebu integrácie do širších digitálnych ekosystémov (obrázok 2.2-2, obrázok 4.1-12).
- **Podceňovanie digitalizácie zo strany konkurentov:** manažéri neberú do úvahy, že aj konkurenči aktívne zavádzajú digitálne technológie, čo môže viesť k strate konkurenčnej výhody.
- **chýbajúca dualita digitalizácie:** generálni riaditelia delegujú zodpovednosť za digitálnu transformáciu na iných vedúcich pracovníkov, čo byrokratizuje kontrolu a spomaľuje proces zmien.

Riešenie týchto výziev si vyžaduje jasné pochopenie a zosúladenie digitálnych stratégií na všetkých

úrovniach organizácie. Pred vytvorením digitálnej stratégie je dôležité pochopiť východiskový bod. Mnohé organizácie majú tendenciu prijímať nové nástroje a platformy bez toho, aby mali úplný prehľad o súčasnom stave.

### **Krok 1: Vykonalte audit svojich súčasných systémov a údajov.**

Pred zmenou procesov je dôležité pochopiť, čo už je zavedené. Vykonanie auditu vám umožní identifikovať nedostatky v správe údajov a pochopiť, aké zdroje je možné využiť. Audit je akýmsi "röntgenom" vašich podnikových procesov. Umožní vám identifikovať rizikové oblasti a určiť, ktoré údaje sú pre váš projekt alebo podnikanie kritické a ktoré sú druhoradé.

Hlavné činnosti:

- Zmapujte svoje IT prostredie (v aplikáciách Draw.io, Lucidchart, Miro, Visio alebo Canva). Uveďte zoznam používanych systémov (ERP, CAD, CAFM, CPM, SCM a iné), ktoré sa podielajú na vašich procesoch a o ktorých sme hovorili v kapitole "Technológie a systémy riadenia v modernej výstavbe" (obr. 1.2-4)
- Posúdte otázky kvality údajov pre každý systém z hľadiska frekvencie duplikátov, možných chýbajúcich hodnôt a nekonzistentnosti formátu v každom systéme.
- Identifikujte "boľavé miesta" - miesta, kde sa procesy môžu pokaziť alebo si často vyžadujú manuálny zásah - import, export a ďalšie procesy validácie.

Ak chcete, aby tím dôveroval správam, musíte sa od začiatku uistíť, že údaje sú správne.

Audit kvality údajov vám ukáže, aké údaje:

- Potrebuje ďalší vývoj (je potrebné nastaviť automatické čistiace procesy alebo ďalšie transformácie)
- Sú to "odpadky", ktoré len upchávajú systémy a možno sa ich zbaviť tak, že sa už nebudú používať v procesoch.

Takýto audit je možné vykonať aj samostatne. Niekedy je však užitočné zapojiť externého konzultanta - najmä z iných odvetví: nový pohľad a nezávislosť od stavebných "zvláštností" pomôžu tiezvo posúdiť súčasný stav a vyhnúť sa typickým nástrahám zaujatosti voči určitým riešeniam a technológiám.

### **Krok 2: Identifikácia kľúčových noriem pre harmonizáciu údajov.**

Po audite je potrebné vytvoriť spoločné pravidlá pre prácu s údajmi. Ako sme uviedli v kapitole "Štandardy: Od náhodných súborov k inteligentnému dátovému modelu", pomôže to odstrániť izolované dátové toky.

Bez jednotného štandardu bude každý tím naďalej pracovať "po svojom" a vy budete udržiavať "zoologickú záhradu" integrácií, kde sa pri každej konverzii strácajú údaje.

#### Hlavné činnosti:

- Vyberte dátové štandardy na výmenu informácií medzi systémami:
  - V prípade tabuľkových údajov to môžu byť štruktúrované formáty, ako napríklad CSV, XLSX alebo efektívnejšie formáty, ako napríklad Parquet
  - Na výmenu voľne štruktúrovaných údajov a dokumentov: JSON alebo XML
- Ovládanie práce s dátovými modelmi:
  - Začnite parametrizáciou úloh na úrovni koncepčného modelu údajov - ako je opísané v kapitole "Modelovanie údajov: koncepčný, logický a fyzický model" (obr. 4.3-2).
  - Pri hlbšom prenikaní do logiky obchodného procesu prejdite k formalizácii požiadaviek pomocou parametrov v logickom a fyzickom modeli (obrázok 4.3-6).
  - Identifikujte kľúčové entity, ich atribúty a vzťahy v rámci procesov a vizualizujte tieto vzťahy - medzi entitami aj medzi parametrami (obrázok 4.3-7).
- Na overovanie a štandardizáciu údajov používajte regulárne výrazy (RegEx) (obrázok 4.4-7), ako sme uviedli v kapitole "Štruktúrované požiadavky a regulárne výrazy RegEx". RegEx nie je zložitá, ale mimoriadne dôležitá téma pri práci na tvorbe požiadaviek na úrovni fyzických dátových modelov.

Bez štandardov na úrovni údajov a vizualizácie procesov nie je možné vytvoriť konzistentné a škálovateľné digitálne prostredie. Pamäťajte, že "zlé údaje sú drahé". A náklady na chyby sa zvyšujú s tým, ako sa projekt alebo organizácia stáva zložitejšou. Zjednotenie formátov, definovanie pomenovania, štruktúry a pravidiel validácie je investíciou do stability a škálovateľnosti budúcich riešení.

#### Krok 3. Implementácia DataOps a automatizácia procesov.

Bez dobre definovanej architektúry budú spoločnosti nevyhnutne čeliť rôznorodým údajom obsiahnutým v oddelených informačných systémoch. Údaje budú neintegrované, duplikované na viacerých miestach a ich údržba bude nákladná.

Predstavte si, že údaje sú voda a dátová architektúra je komplexný systém potrubí, ktorý túto vodu prepravuje od zdroja jej uskladnenia až po miesto jej použitia. Práve dátová architektúra určuje, ako sa informácie zhromažďujú, ukladajú, transformujú, analyzujú a dodávajú koncovým používateľom alebo aplikáciám.

DataOps (Data Operations) je metodika, ktorá integruje zber, čistenie, overovanie a používanie údajov do jedného automatizovaného procesu, ako sme podrobne rozobrali v 8. časti knihy.

Hlavné činnosti:

- Vytváranie a prispôsobovanie ETL -konvektory na automatizáciu procesov:
  - Výpis: organizujte automatický zber údajov z dokumentov PDF (obr. 4.1-2, obr. 4.1-5, obr. 4.1-7), tabuľiek Excel, modelov CAD (obr. 7.2-4), systémov ERP a ďalších zdrojov, s ktorými pracujete
  - Transformácia: nastavenie automatických procesov na transformáciu údajov do jednotného štruktúrovaného formátu a automatizáciu výpočtov, ktoré budú prebiehať mimo uzavretých aplikácií (obr. 7.2-8).
  - Načítanie: skúste vytvoriť automatické odosielanie údajov do súhrnných tabuľiek, dokumentov alebo centralizovaných úložísk (obr. 7.2-9, obr. 7.2-13, obr. 7.2-16).
- Automatizujte procesy výpočtu a QTO (Quantity Take-Off), ako sme o tom hovorili v kapitole "QTO Quantity Take-Off: Zoskupovanie údajov o projekte podľa atribútov":
  - Nastavte automatickú extrakciu objemov z modelov CAD pomocou rozhraní API, zásuvných modulov alebo nástrojov reverzného inžinierstva (obr. 5.2-5).
  - Vytvorenie pravidiel na zoskupovanie prvkov pre rôzne triedy podľa atribútov vo forme tabuľiek (obrázok 5.2-12)
  - Pokúste sa automatizovať často sa opakujúce výpočty objemu a nákladov mimo modulárnych uzavretých systémov (obr. 5.2-15).
- Začnite používať Python a Pandas na spracovanie údajov, ako sme o tom hovorili v kapitole "Python Pandas: nenahraditeľný nástroj na prácu s údajmi":
  - Aplikovať DataFrame na prácu so súbormi XLSX a automatizovať spracovanie tabuľkových údajov (obr. 3.4-6)
  - Automatizácia agregácie a transformácie informácií prostredníctvom rôznych knižníc Pythonu
  - Používanie LLM na zjednodušenie písania hotových blokov kódu a celých Potrubia (obrázok 7.2-18).
  - Skúste si v jazyku Python zostaviť potrubnú linku, ktorá nájde chyby alebo zaznamená anomálie a pošle oznámenie zodpovednej osobe (napr. vedúcemu projektu) (obrázok 7.4-2).

Automatizácia založená na princípoch DataOps vám umožní prejsť od manuálneho a fragmentovaného spracovania údajov k udržateľnému a opakovateľnému procesom. Tým sa nielen zníži záťaž zamestnancov, ktorí sa každý deň zaoberajú rovnakými transformáciami, ale zároveň sa výrazne zvýši spoľahlivosť, škálovateľnosť a transparentnosť celého informačného systému.

#### Krok 4: Vytvorenie otvoreného ekosystému správy údajov.

Napriek vývoju uzavretých modulárnych systémov a ich integrácií s novými nástrojmi čelia spoločnosti vážnemu problému - rastúca zložitosť takýchto systémov prevyšuje ich užitočnosť. Pôvodná myšlienka

vytvoriť jedinú vlastnú platformu pokrývajúcu všetky podnikové procesy viedla k nadmernej centralizácii, kde si akékoľvek zmeny vyžadujú značné zdroje a čas na prispôsobenie.

Ako sme uviedli v kapitole "Firemné mycélium: Ako údaje spájajú podnikové procesy", efektívna správa údajov si vyžaduje otvorený a jednotný ekosystém, ktorý spája všetky zdroje informácií.

Kľúčové prvky ekosystému:

■ Vyberte vhodné dátové úložisko:

- Pre tabuľky a výpočty použite databázy - napríklad PostgreSQL alebo MySQL (obr. 3.1-7).
- Pre dokumenty a správy môže byť vhodné cloudové úložisko (Disk Google, OneDrive) alebo systémy, ktoré podporujú formát JSON.
- Pozrite si možnosti dátových skladov, dátových jazier a ďalších nástrojov na centralizované ukladanie a analýzu veľkého množstva informácií (obrázok 8.1-8).

■ Implementácia riešení na prístup k vlastným údajom:

- Ak používate proprietárne systémy, nakonfigurujte prístup k nim prostredníctvom API alebo SDK, aby ste získali údaje na externé spracovanie (obr. 4.1-2).
- Zoznámte sa s možnosťami nástrojov reverzného inžinierstva pre formáty CAD (obrázok 4.1-13).
- Nastavte ETL-pipeline, ktoré pravidelne zbierajú údaje z aplikácií alebo serverov, konvertujú ich do otvorených štruktúrovaných formátov a ukladajú do úložísk (obrázok 7.2-3).
- diskutovať v rámci tímu o tom, ako zabezpečiť prístup k údajom bez potreby proprietárneho softvéru
- Pamäťajte: údaje sú dôležitejšie ako rozhrania. Dlhodobú hodnotu poskytuje štruktúra a dostupnosť informácií, nie konkrétné nástroje používateľského rozhrania.

■ Premýšľajte o vytvorení centra excelentnosti (CoE) pre údaje, o ktorom sme hovorili v kapitole "Centrum excelentnosti (CoE) pre modelovanie údajov", alebo o tom, ako môžete poskytovať odborné znalosti o údajoch iným spôsobom (obrázok 4.3-9).

Ekosystém správy údajov vytvára jednotný informačný priestor, v ktorom všetci účastníci projektu pracujú s konzistentnými, aktuálnymi a overenými informáciami. Je základom pre škálovateľné, flexibilné a spoľahlivé digitálne procesy

## Uvoľnenie potenciálu údajov: 5-10 krokov k digitálnej zrelosti

Okrem technickej integrácie je dôležitým faktorom úspešnej implementácie digitálnych riešení ich prijatie koncovými používateľmi. Zapojenie zákazníkov alebo používateľov do merania výkonnosti je výzvou na zlepšenie používateľskej skúsenosti aj na riadenie zmien v spoločnosti. Ak riešenie nezapadá do známeho pracovného postupu alebo nerieši skutočné problémy používateľov alebo zákazníkov, nebude sa používať a žiadne dodatočné opatrenia a stimuly to nespravia.

Transformácia je iteračný proces založený na analýze údajov o interakcií používateľov s novými procesmi, s častými testovacími cyklami, neustálou spätnou väzbou a vylepšovaním.

### **Krok 5: Vybudujte kultúru údajov, zaškoľujte zamestnancov a získavajte spätnú väzbu**

Ani ten najdokonalejší systém nebude fungovať bez zapojenia zamestnancov. Musíte vytvoriť prostredie, v ktorom sa údaje používajú na dennej báze a tím chápe ich hodnotu.

V verejnenej správe vlády Spojeného kráľovstva do roku 2024 "Analýza údajov a umelá inteligencia pri realizácii vládnych projektov" sa uvádzá [83], že pre úspešnú implementáciu analýzy údajov a umelej inteligencie je rozhodujúca odborná príprava odborníkov s potrebnými kompetenciami v oblasti spracovania a interpretácie údajov.

Nedostatok odborných znalostí v oblasti analýzy údajov je jedným z kľúčových problémov, ktoré obmedzujú digitálnu transformáciu. Vedúci pracovníci sú zvyknutí na zaužívané postupy: štvrtročné cykly, prioritné iniciatívy a tradičné spôsoby napredovania projektov. Zmena si vyžaduje špeciálneho lídra - dostatočne vysoko postaveného, aby mal vplyv, ale nie tak vysoko, aby mal čas a motiváciu viesť dlhodobý transformačný projekt.

Hlavné činnosti:

- Uvedomenie si potreby prechodu od subjektívnych rozhodnutí založených na názore vysoko plateného zamestnanca (HiPPO) ku kultúre rozhodovania založenej na faktoch a údajoch, ako sa uvádzá v kapitole "HiPPO alebo nebezpečenstvo názoru pri rozhodovaní" (obrázok 2.1-9).
- Organizujte systematické školenia:
  - Usporiadať školenia o tom, ako používať štruktúrované údaje, a pozvať odborníkov z iných odvetví, ktorí nie sú zaujatí produktmi a koncepciami populárnymi v súčasnom stavebnictve.
  - Diskutovať s kolegami o prístupoch a nástrojoch na analýzu údajov a samostatne sa naučiť praktickú prácu s nástrojmi, ako sú Python, pandas a LLM (obr. 4.1-3, obr. 4.1-6)
  - Vytvoriť knižnicu výukových materiálov (najlepšie s krátkymi videami) na tému štruktúrovania údajov (obr. 3.2-15) a tvorby dátových modelov (obr. 4.3-6, obr. 4.3-7).
- Využívajte moderné vzdelenacie technológie:
  - Používanie jazykových modelov (LLM) na podporu manipulácie s kódom a údajmi vrátane generovania kódu, refaktorovania a analýzy, ako aj spracovania a interpretácie tabuľkových informácií (obrázok 3.4-1)
  - Preskúmajte, ako sa dá kód vytvorený pomocou LLM prispôsobiť a integrovať do kompletného riešenia Pipeline pri práci v offline vývojovom prostredí (IDE) (obr. 4.4-14, obr. 5.2-13).

Ak manažér naďalej rozhoduje "staromódnym spôsobom", žiadne školenie nepresvedčí ľudí, aby brali analytiku vážne.

Budovanie dátovej kultúry nie je možné bez neustálej spätej väzby. Spätná väzba umožňuje identifikovať nedostatky v procesoch, nástrojoch a stratégiah, ktoré nie je možné odhaliť prostredníctvom interných správ alebo formálnych KPI metrík. Pochvalné pripomienky používateľov vašich riešení neprinesú praktickú hodnotu. Cenná je kritická spätná väzba, najmä ak je založená na konkrétnych pozorovaniach a faktoch. Získanie takýchto informácií si však vyžaduje úsilie: musíte vytvoriť procesy, v ktorých sa účastníci - interní aj externí - môžu podeliť o pripomienky (môže mať zmysel robiť to anonymne) bez skreslenia a bez obáv, že ich názory môžu ovplyvniť ich vlastnú prácu. Je dôležité, aby tak robili bez skreslenia a bez obáv z negatívnych dôsledkov pre seba.

*Každé učenie je v konečnom dôsledku samoučením [165].*

- Milton Friedman, americký ekonóm a štatistik

Zavedenie analytickej kultúry by malo byť sprevádzané pravidelným overovaním ich účinnosti v praxi (návratnosť investícii, kľúčové ukazovatele výkonnosti), čo možno dosiahnuť len prostredníctvom štruktúrovanej spätej väzby od zamestnancov, zákazníkov a partnerov. To umožňuje spoločnostiam nielen vyhnúť sa opakovaniu chýb, ale aj rýchlejšie sa prispôsobiť zmenám v prostredí. Existencia mechanizmu na zber a analýzu spätej väzby je jedným zo znakov vyspelosti organizácie, ktorá prechádza od príležitostných digitálnych iniciatív k udržateľnému modelu neustáleho zlepšovania (obrázok 2.2-5).

## Krok 6: Od pilotných projektov k rozšíreniu

*Vyberajte si dostatočne veľké bitky, aby na nich záležalo, a dostatočne malé, aby ste ich vyhrali.*

- Jonathan Kozol

Spustenie digitálnej transformácie "naraz a všade" je veľmi riskantné. Efektívnejší prístup je začať s pilotnými projektmi a postupne rozširovať úspešné postupy.

Hlavné činnosti:

- Vyberte správny projekt pre pilotný projekt:
  - Definujte konkrétny obchodný cieľ alebo proces s merateľnými výsledkami (KPI, ROI)

(obr. 7.1-5).

- Vyberte proces automatizácie ETL, napríklad automatickú validáciu údajov alebo výpočet pracovného zaťaženia (QTO) pomocou programov Python a Pandas (obrázok 5.2-10)
- Stanovte jasné metriky úspechu (napr. skrátiť čas na vypracovanie inšpekčných specifikácií alebo správ o overovaní údajov z týždňa na jeden deň).
- Využívajte opakované prístupy:
  - Začnite s jednoduchými procesmi konverzie údajov a vytvorte prúdové konverzie viacformátových údajov do formátov, ktoré potrebujete pre svoje procesy (obr. 4.1-2, obr. 4.1-5).
  - Postupne zvyšujte zložitosť úloh a rozširujte automatizáciu procesov vytváraním kompletnej Potrubia v IDE na základe zdokumentovaných blokov kódu (obr. 4.1-7, obr. 7.2-18).
  - zdokumentovať a zaznamenať (najlepšie pomocou krátkych videí) úspešné riešenia a zdieľať ich s kolegami alebo v odborných komunitách.
- Vypracovať šablóny a sprievodnú dokumentáciu na replikovanie takýchto riešení, aby ich mohli efektívne používať vaši kolegovia (alebo členovia odbornej komunity vrátane používateľov sociálnych médií).

Postupný "roll-up" umožňuje zachovať vysokú kvalitu zmien a neupadnúť do chaosu paralelných implementácií. Stratégia "od malého k veľkému" minimalizuje riziká a umožňuje vám poučiť sa z malých chýb bez toho, aby prerástli do kritických problémov.

Prechod od projektového prístupu, pri ktorom sú zamestnanci zapojení len čiastočne, k vytváaniu stálych tímov (napr. odborných centier - CoE) pomáha zabezpečiť udržateľný vývoj produktu aj po vydaní jeho prvej verzie. Takéto tímy nielen podporujú existujúce riešenia, ale ich aj nadálej zlepšujú.

Znižuje sa tak závislosť od zdĺhavého schvaľovania: členovia tímu sú oprávnení prijímať rozhodnutia v rámci svojich kompetencií. Výsledkom je, že manažéri sú oslobodení od potreby mikromanažmentu a tímy sa môžu sústrediť na vytváranie skutočnej hodnoty.

Vývoj nových riešení nie je šprint, ale maratón. Úspešní sú tí, ktorí sa na začiatku zameriavajú na dlhodobú a sústavnú prácu.

Je dôležité uvedomiť si, že technológia si vyžaduje neustály vývoj. Investície do dlhodobého vývoja technologických riešení sú základom úspešnej prevádzky.

### Krok 7: Používajte otvorené formáty údajov a riešenia

Ako sme uviedli v kapitolách o modulárnych platformách (ERP, PMIS, CAFM, CDE atď.), je dôležité zamerať sa na otvorené a univerzálné formáty údajov, ktoré zabezpečia nezávislosť od riešení dodávateľa a zvýšia dostupnosť informácií pre všetky zainteresované strany.

## Hlavné činnosti:

- Prechod z uzavretých formátov na otvorené formáty:
  - Používajte otvorené formáty namiesto proprietárnych formátov alebo nájdite spôsob, ako nastaviť automatické nahrávanie alebo konverziu uzavretých formátov na otvorené formáty (obrázok 3.2-15).
  - Implementácia nástrojov na prácu s formátm Parquet, CSV, JSON, XLSX, ktoré predstavujú štandardy výmeny medzi väčšinou moderných systémov (obr. 8.1-2).
  - Ak práca s 3D geometriou zohráva vo vašich procesoch dôležitú úlohu, zvážte používanie otvorených formátov, ako sú USD, glTF, DAE alebo OBJ (obrázok 3.1-14).
- Používanie vektorových databáz databáz na efektívnu analýzu a vyhľadávanie informácií:
  - Použite Bounding Box a iné metódy na zjednodušenie 3D -geometrie (obrázok 8.2-1).
  - Premyslite si, kde môžete implementovať vektorizáciu údajov - prevod textov, objektov alebo dokumentov na číselné reprezentácie (obrázok 8.2-2).
- Aplikujte nástroje na analýzu veľkých objemov údajov:
  - Organizovať ukladanie nahromadených historických údajov (napr. PDF, XLSX, CAD) do formátov vhodných na analýzu (Apache Parquet, CSV, ORC) (obr. 8.1-2).
  - Začnite používať základné štatistické metódy a pracovať s reprezentatívnymi vzorkami - alebo sa aspoň oboznámte so základnými princípmi štatistiky (obrázok 9.2-5).
  - Implementovať a naučiť sa nástroje na vizualizáciu údajov a vzťahov medzi údajmi s cieľom vizualizovať výsledky analýzy. Bez dobrej vizualizácie nie je možné úplne pochopiť samotné údaje ani procesy na nich založené (obrázok 7.1-4).

Prechod na otvorené formáty údajov a zavedenie nástrojov na analýzu, ukladanie a vizualizáciu informácií vytvára základ pre udržateľnú a nezávislú digitálnu správu. Tým sa nielen znižuje závislosť od dodávateľov, ale zároveň sa zabezpečuje rovnaký prístup k údajom pre všetky zainteresované strany.

## Krok 8. Začnite implementovať strojové učenie na predpovedanie

Mnohé spoločnosti nahromadili obrovské množstvo údajov - aké sú "informačné gejzíry", ktoré sú stále nevyužité. Tieto údaje boli zozbierané v stovkách alebo tisícach projektov, ale v ďalších procesoch boli často použité len raz alebo vôbec. Dokumenty a modely uložené v uzavretých formátoch a systémoch sú často vnímané ako zastarané a nepoužiteľné. V skutočnosti sú však cenným zdrojom - základom pre analýzu urobených chýb, automatizáciu rutinných operácií a vývoj inovatívnych riešení pre automatickú klasifikáciu a rozpoznávanie prvkov v budúcich projektoch.

Klúčovou výzvou je naučiť sa tieto údaje získavať a transformovať ich na využiteľné poznatky. Ako sa uvádzajú v kapitole o strojovom učení a predpovediach, techniky strojového učenia majú potenciál výrazne zlepšiť presnosť odhadov a predpovedí v rôznych procesoch súvisiacich so stavebníctvom.

Plné využitie nahromadených údajov otvára cestu k zvýšeniu efektívnosti, zníženiu rizika a budovaniu udržateľných digitálnych procesov.

Hlavné činnosti:

- Začnite s jednoduchými algoritmami:
  - Skúste použiť lineárnu regresiu - s využitím nápovedy z LLM - na predpovedanie rekurentného výkonu v súboroch údajov, kde závislosti od veľkého počtu faktorov chýbajú alebo sú minimálne (obrázok 9.3-4).
  - Zvážte, v ktorých fázach vašich procesov by sa teoreticky mohol uplatniť algoritmus k-najbližších susedov (k-NN) - napríklad pri klasifikačných úlohách, posudzovanie podobnosti objektov alebo predpovedaní na základe historických analógov (obr. 9.3-5).
- Zhromažďovanie a štruktúrovanie údajov na trénovanie modelov:
  - Zhromažďovanie historických údajov o projekte na jednom mieste a v jednom formáte (obrázok 9.1-10)
  - Práca na kvalite a reprezentatívnosti cvičných vzoriek prostredníctvom automatizovaného ETL (obrázok 9.2-8).
  - Naučte sa rozdeliť údaje na trénovaciu a testovaciu množinu, ako sme to urobili v príklade s dátovou množinou Titanic (obrázok 9.2-6, obrázok 9.2-7).
- Zvážte rozšírenie použitia techník strojového učenia na širokú škálu úloh, od predpovedania časového harmonogramu projektu až po optimalizáciu logistiky, riadenie zdrojov a včasné identifikáciu potenciálnych problémov.

Strojové učenie je nástroj na premenu archivovaných údajov na cenný zdroj na predpovedanie, optimalizáciu a informované rozhodovanie. Začnite s malými súbormi údajov (obrázok 9.2-5) a jednoduchými modelmi a postupne zvyšujte ich zložitosť.

### Krok 9. Integrácia internetu vecí a moderných technológií zberu údajov

Svet stavebníctva sa rýchlo stáva digitálnym: každá stavebná fotografia, každý príspevok v Teams je už súčasťou väčšieho procesu parametrizácie a symbolizácie reality. Tak ako GPS kedysi zmenilo logistiku, internet vecí, RFID a technológie automatického zberu údajov menia stavebníctvo. Ako sa uvádzajú v kapitole "Internet vecí a inteligentné zmluvy", digitálne stavenisko so senzormi a automatickým monitorovaním je budúcnosťou tohto odvetvia.

Hlavné činnosti:

- Implementácia IoT -zariadení, RFID -tagov a podrobnejšie informácie o procesoch s nimi spojených:
  - Vyhodnotiť, ktoré oblasti alebo fázy projektu môžu mať najväčší prínos z návratnosti investícií (ROI) - napríklad monitorovanie teploty, vibrácií, vlhkosti alebo pohybu.
  - Zvážte použitie RFID na sledovanie materiálov, nástrojov a zariadení v celom dodávateľskom reťazci.

- Zvážte, ako možno zhromaždené údaje integrovať do jednotného informačného systému, napríklad Apache NiFi, na automatizované spracovanie a analýzu v reálnom čase (obrázok 7.4-5).

■ Zaviesť systém monitorovania v reálnom čase:

- Vytvárať informačné panely na sledovanie kľúčových procesných alebo projektových ukazovateľov pomocou vizualizačných nástrojov, ako sú Streamlit, Flask alebo Power BI)
- Nastavenie automatických oznámení na signalizáciu kritických odchýlok od plánu alebo noriem (obr. 7.4-2).
- Posúdiť potenciál prediktívnej údržby zariadení na základe zozbieraných údajov a zistených modelov (obrázok 9.3-6).

■ Kombinujte údaje z rôznych zdrojov:

- Začnite vizualizáciou dátového modelu na fyzickej úrovni - zohľadnite štruktúru informačných tokov a kľúčových parametrov pochádzajúcich zo systémov CAD, zariadení IoT a platforem ERP (obr. 4.3-1).
- Začnite vytvorením hrubého náčrtu jednotnej platformy na analýzu údajov a podporu rozhodovania manažmentu. Zachyťte kľúčové funkcie, zdroje údajov, používateľov a predpokladané scenáre použitia (obrázok 4.3-7).

Čím skôr začnete prepájať skutočné procesy s digitálnym svetom, tým skôr ich budete môcť riadiť pomocou údajov - efektívne, transparentne a v reálnom čase.

### Krok 10. Pripravte sa na budúce zmeny v odvetví

Stavebné spoločnosti sú pod neustálym tlakom vonkajšieho prostredia: hospodárske krízy, technologické skoky, regulačné zmeny. Podobne ako les, ktorý musí odolávať dažďu, snehu, suchu a páliacemu slnku, aj spoločnosti žijú v podmienkach neustálej adaptácie. A tak ako sa stromy vďaka hlbokému koreňovému systému stávajú odolnými voči mrazu a suchu, životaschopné a konkurencieschopné zostávajú len tie organizácie, ktoré majú pevný základ automatizovaných procesov, schopnosť predvídať zmeny a pružne prispôsobovať strategie.

Ako bolo uvedené v kapitole "Stratégie prežitia: budovanie konkurenčných výhod", stavebníctvo vstupuje do fázy radikálnej transformácie. Interakcia medzi objednávateľom a zhотовiteľom sa posúva k modelu presvedčania, kde transparentnosť, predvídateľnosť a digitálne nástroje nahradzajú tradičné prístupy. V tejto novej realite nevyhľávajú tí najväčší, ale tí najflexibilnejší a technologicky najvyspelejší.

Hlavné činnosti:

■ Analyzujte zraniteľnosť podnikov v kontexte otvorených údajov:

- Posúdte, ako by demokratizácia prístupu k údajom v rámci uberizácie mohla mať zničujúci vplyv na vašu konkurenčnú výhodu a vaše podnikanie (obrázok 10.1-5).
- Premyslite si stratégii prechodu od neprehľadných a izolovaných procesov k

obchodným modelom založeným na otvorených riešeniach, interoperabilite systémov a transparentnosti údajov (obr. 2.2-5).

■ Vypracujte dlhodobú digitálnu stratégiu:

- Rozhodnite sa, či chcete byť inovačným lídom, alebo uprednostňujete scenár "dobiehania", v ktorom budete šetriť svoje zdroje.
- Opíste etapy: krátkodobá (automatizácia procesov, štruktúrovanie údajov), strednodobá (implementácia LLM a ETL), dlhodobá (digitálne ekosystémy, centralizované úložiská).

■ Premýšľajte o rozšírení svojho portfólia služieb:

- Zväžiť ponuku nových služieb (zameraných na energetickú účinnosť, ESG, dátové služby). O nových obchodných modeloch budeme viac hovoriť v nasledujúcej kapitole
- Snažte sa vystupovať ako spoľahlivý technologický partner, ktorý podporuje celý životný cyklus zariadenia, od návrhu až po prevádzku. Dôvera vo vás by mala byť založená na systematickom prístupe, transparentných procesoch a schopnosti poskytovať udržateľné technologické riešenia

V transformačnom prostredí nevyhrávajú tí, ktorí na zmeny len reagujú, ale tí, ktorí sú aktívni. Flexibilita, otvorenosť a digitálna vyspelosť sú základom udržateľnosti pri budovaní zajtrajška.

## Plán transformácie: od chaosu k spoločnosti riadenej údajmi

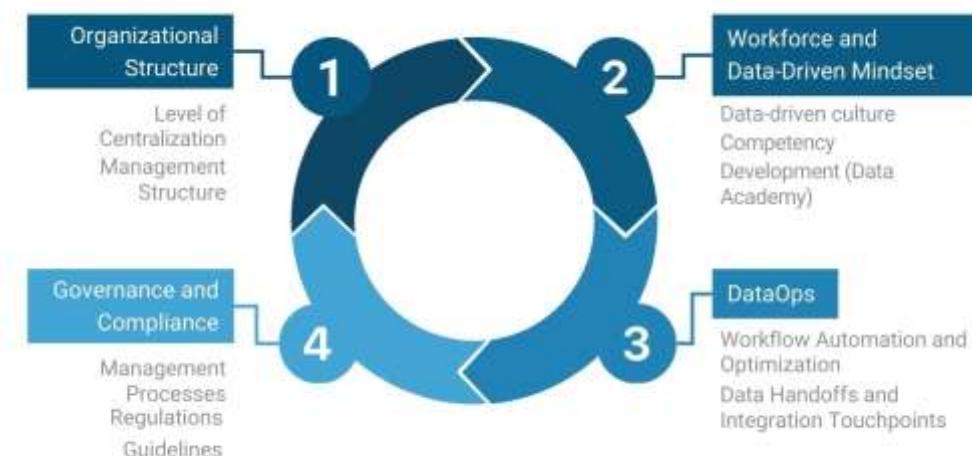
Nasledujúci plán môže slúžiť ako počiatočné kritérium - východiskový bod pre formovanie vašej vlastnej stratégie digitálnej transformácie založenej na údajoch:

- **Audit a normy:** analýza súčasného stavu, zjednotenie údajov
- **Štruktúrovanie a klasifikácia údajov:** automatizácia transformácie neštruktúrovaných a voľne štruktúrovaných údajov
- **Automatizácia zoskupení,** výpočtov a kalkulácií: používanie nástrojov a knižníc s otvoreným zdrojovým kódom na automatizáciu
- **Ekosystém a COE:** vytvorenie interného tímu na vytvorenie jednotného dátového ekosystému v spoločnosti
- **Kultúra a vzdelávanie:** prechod od riešení HiPPO k riešeniam založeným na údajoch
- **Pilotné projekty, spätná väzba a rozširovanie:** postupujte iteratívne: testujte nové metódy v obmedzenom rozsahu, získavajte platnú spätnú väzbu a postupne riešenia rozširujte.
- **Otvorené formáty:** používanie univerzálnych a otvorených formátov pre nezávislosť od dodávateľov softvéru
- **Strojové učenie:** zabudovanie algoritmov ML do procesov na predpovedanie a optimalizáciu
- **IoT a digitálne stavenisko:** integrácia moderných technológií zberu údajov do procesov

### ■ Strategická adaptácia: príprava na budúce zmeny v odvetví

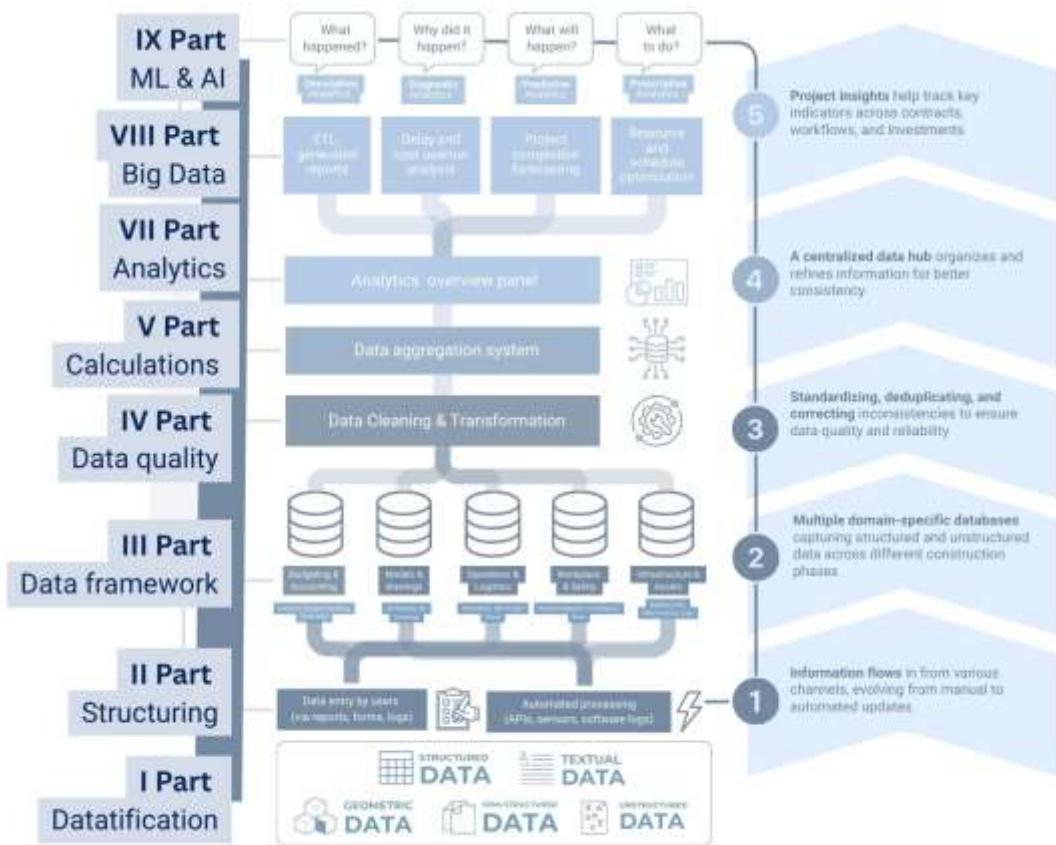
Najdôležitejšie je pamätať na to, že "samotné údaje spoločnosť nezmenia: zmenia ju ľudia, ktorí s nimi vedia pracovať". Dôraz na kultúru, transparentné procesy a záväzok neustáleho zlepšovania

Systémový prístup umožňuje prechod od izolovaných digitálnych iniciatív k plnohodnotnému modelu riadenia založeného na údajoch, v ktorom sa rozhodnutia nezakladajú na intuícii alebo predpokladoch, ale na údajoch, faktoch a matematicky vypočítaných pravdepodobnostiach. Digitálna transformácia stavebníctva nie je len o osvojení si technológií, ale o formovaní obchodného ekosystému, v ktorom sa informácie o projektoch prenášajú bez problémov a iteratívne medzi rôznymi systémami. Algoritmy strojového učenia pritom zabezpečujú automatickú, priebežnú analýzu, prognózovanie a optimalizáciu procesov. V takomto prostredí sa špekulácie a skryté údaje stávajú irrelevantnými - zostávajú len overené modely, transparentné výpočty a predvídateľné výsledky.



Obr. 10.2-3 Klúčové prvky úspešného riadenia údajov na úrovni podniku.

Každá časť knihy zodpovedá určitej fáze spracovania a analýzy údajov v stavebných projektoch (obr. 2.2-5). Ak sa chcete vrátiť k niektornej z už preberaných tém a pozrieť sa na ňu z hľadiska komplexného chápania toku využívania údajov, môžete sa odvolať na názvy častí na obr. 10.2-4.



Obr. 10.2-4 Časti knihy v kontexte potrubia spracovania údajov (obr. 2.2-5): od digitalizácie informácií po analytiku a umelú inteligenciu.

Bez ohľadu na veľkosť vašej organizácie, úroveň technologickej vyspelosti alebo rozpočet môžete začať s prechodom na prístup založený na údajoch už dnes. Aj malé kroky správnym smerom prinesú časom výsledky.

Transformácia založená na údajoch nie je jednorazový projekt, ale nepretržitý, opakujúci sa proces zlepšovania, ktorý zahŕňa zavádzanie nových nástrojov, prepracovanie procesov a rozvoj kultúry rozhodovania založeného na údajoch.

## Budovanie v priemysle 5.0: Ako zarábať peniaze, keď sa už nemôžete skrývať

Stavebné spoločnosti už dlho zarábajú na neprehľadnosti procesov. Hlavným obchodným modelom boli špekulácie - nadhodnocovanie nákladov na materiál, rozsah prác a percentuálne prirážky v uzavretých ERP - a PMIS - systémoch, ktoré sú neprístupné externým auditom. Obmedzený prístup zákazníkov a ich dôveryhodných osôb k pôvodným údajom o projekte vytvoril pôdu pre schémy, v ktorých bolo takmer nemožné overiť spoľahlivosť výpočtov.

Tento model však rýchlo stráca význam. S demokratizáciou prístupu k údajom, vznikom LLM, príchodom otvorených údajov a nástrojov na automatizáciu ETL prechádza odvetvie na nový štandard práce.

V dôsledku toho už netransparentnosť nie je konkurenčnou výhodou - čoskoro sa stane bremenom, s ktorým bude ľahšie sa rozlúčiť. Transparentnosť sa z možnosti zmení na podmienku udržania sa na trhu.

S kym budú v novej digitálnej realite spolupracovať zákazníci - banky, investori, fyzickí zákazníci, súkromný kapitál, zákazníci štátnej správy? Odpoveď je zrejmá: s tými, ktorí sú schopní poskytnúť nielen výsledok, ale aj zdôvodnenie každého kroku na ceste k nemu. S rastúcim objemom otvorených údajov si partneri a zákazníci budú vyberať spoločnosti, ktoré zaručujú transparentnosť, presnosť a predvídateľnosť výsledkov.

V tejto súvislosti vznikajú nové obchodné modely, ktoré sú založené skôr na správe údajov a dôvere než na špekuláciách:

■ **Predaj procesov namiesto metrov štvorcových:** kľúčovým prínosom sa stáva dôvera a efektívnosť namiesto konkrétnych zmlúv so zľavou. Hlavnou hodnotou bude predvídateľnosť výsledku na základe spoľahlivých a overených údajov. Moderné spoločnosti nebudú predávať stavebný objekt ako taký, ale:

- presné termíny a transparentné pracovné harmonogramy;
- primerané odhady, podložené výpočtami;
- úplná digitálna sledovateľnosť a kontrola vo všetkých fázach projektu.

■ **Inžinierstvo a analyтика ako služba:** model "Data-as-a-Service" (spôsob poskytovania hotových údajov používateľom prostredníctvom internetu ako služby), kde sa každý projekt stáva súčasťou digitálneho dátového reťazca a obchodná hodnota spočíva v schopnosti riadiť tento reťazec. Spoločnosti sa transformujú na inteligentné platformy ponúkajúce riešenia založené na automatizácii a analytike:

- Automatizovaná a transparentná príprava odhadov a plánov;
- hodnotenie rizika a načasovania na základe algoritmov strojového učenia;
- výpočet environmentálnych ukazovateľov (ESG, CO<sub>2</sub>, energetická účinnosť);
- generovanie správ z auditovaných otvorených zdrojov.

■ **Produktivizácia inžinierskych znalostí:** vývoj spoločnosti sa môže opakovane používať v rámci spoločnosti a distribuovať ako samostatný produkt - tvorí dodatočný zdroj príjmov prostredníctvom digitálnych služieb. V novom prostredí spoločnosti vytvárajú nielen projekty, ale aj digitálne aktíva:

- knižnice komponentov a šablóny odhadov;
- automatizované overovacie moduly;
- Open-source pluginy a skripty (predaj konzultácií) na prácu s údajmi.

■ **Nový typ spoločnosti: Data-Driven Integrator:** hráč na trhu, ktorý nie je závislý od konkrétnych dodávateľov softvéru alebo modulárnych systémov a nie je "uzamknutý" v jednom softvérovom rozhraní. S údajmi operuje voľne - a na nich buduje svoju konkurencieschopnosť. Stavebná spoločnosť budúcnosti nie je len dodávateľom, ale informačným integrátorom, ktorý je schopný vykonávať pre zákazníka tieto funkcie:

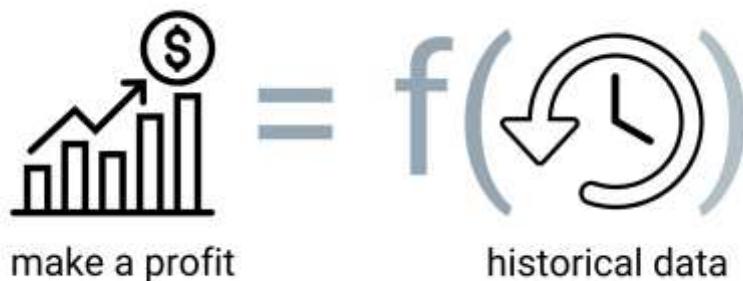
- Kombinujte údaje z rôznych zdrojov a vykonávajte analýzy;
- Zabezpečenie transparentnosti a dôveryhodnosti procesov;
- poradenstvo v oblasti optimalizácie obchodných procesov;
- vyvíjať nástroje, ktoré pracujú v ekosystéme otvorených údajov, LLM, ETL a Pipelines.

Priemysel 5.0 (obr. 2.1-12) znamená koniec "éry ručného priemerovania pomerov" a večerných stretnutí generálnych riaditeľov s oddelením odhadov a účtovníctva. Všetko, čo bolo predtým skryté - výpočty, odhady, objemy - sa stáva otvoreným, overiteľným a zrozumiteľným aj pre neodborníkov. Vítazmi budú tí, ktorí sa ako prví preorientujú. Všetci ostatní zostanú mimo novej digitálnej ekonomiky stavebného sektora.

## ZÁVER

Stavebníctvo vstupuje do éry zásadných zmien. Od prvých záznamov na hlinených tabuľkách až po obrovské množstvo digitálnych údajov, ktoré prúdia z projektových serverov a stavenísk, história riadenia informácií v stavebníctve vždy odrážala vyspelosť technológie svojej doby. Dnes, s nástupom automatizácie, otvorených formátov a inteligentných analytických systémov, odvetvie nečelí postupnému vývoju, ale rýchlej digitálnej transformácií.

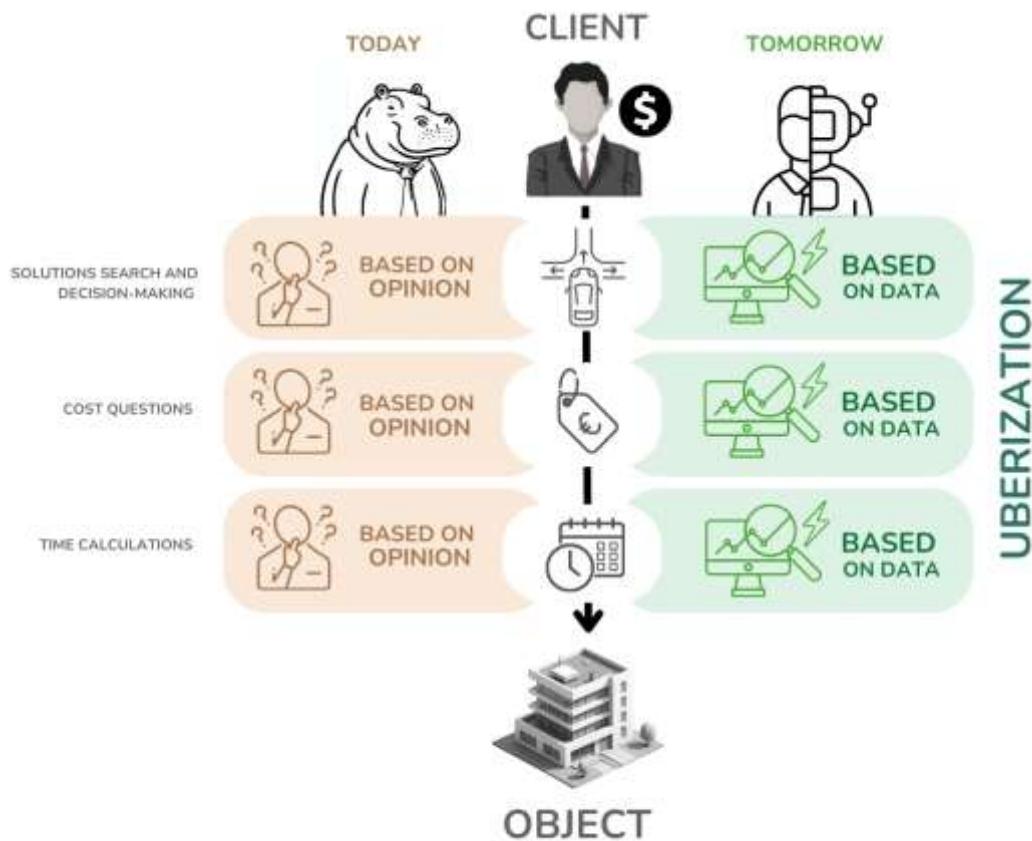
Tak ako v iných odvetviach hospodárstva, aj v stavebníctve bude potrebné prehodnotiť nielen nástroje, ale aj zásady práce. Spoločnosti, ktoré v minulosti diktovali trh a slúžili ako hlavný sprostredkovateľ medzi klientom a projektom, strácajú svoje jedinečné postavenie. Do popredia sa dostáva dôvera a schopnosť pracovať s údajmi - od ich zberu a štruktúrovania až po analýzu, prognózovanie a automatizáciu rozhodnutí.



Obr. 10.2-1 Štruktúrované historické údaje sú palivom pre efektívne a riadené podnikanie.

Táto kniha podrobne opisuje kľúčové princípy správy údajov v stavebníctve - od auditu a štandardizácie až po automatizáciu procesov, používanie vizualizačných nástrojov a implementáciu inteligentných algoritmov. Pozreli sme sa na to, ako aj s obmedzenými združeniami môžete vybudovať funkčnú dátovú architektúru a začať prijímať rozhodenia na základe overiteľných faktov, a nie intuicie. Práca s údajmi už nie je len úlohou IT oddelenia - stáva sa základom kultúry riadenia, ktorá určuje flexibilitu, prispôsobivosť a dlhodobú udržateľnosť spoločnosti.

Aplikácia technológií strojového učenia, systémov automatického spracovania, digitálnych dvojčiat a otvorených formátov už dnes umožňuje eliminovať ľudský faktor tam, kde bol predtým rozhodujúci. Výstavba smeruje k autonómii a kontrolovanosti, pričom pohyb od myšlienky k realizácii projektu možno prirovnáť k navigácii v režime autopilota: bez závislosti od subjektívnych rozhodení, bez potreby manuálneho zásahu v každej fáze, ale s plnou digitálnou sledovateľnosťou a kontrolou (obr. 10.2-2).



Obr. 10.2-2 Posun od rozhodovania na základe názorov dôležitých expertov (HiPPO) k analýze údajov bude podporovať predovšetkým zákazník.

Ak si osvojíte metódy, princípy a nástroje uvedené v tejto knihe, budete môcť vo svojej spoločnosti začať robiť rozhodnutia založené na údajoch, a nie na intuícii. Budete tiež schopní spúštať reťazce modulov v LLM, kopírovať hotové ETL Pipelines do svojho vývojového prostredia (IDE) a automaticky spracovávať údaje, aby ste získali potrebné informácie v požadovanej forme. V nadväznosti na kapitoly knihy o veľkých objemoch údajov a strojovom učení potom budete môcť realizovať zložitejšie scenáre - získavať nové poznatky z historických údajov a aplikovať algoritmy strojového učenia na predpovedanie a optimalizáciu procesov.

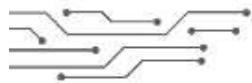
Otvorené údaje a procesy poskytnú základ pre presnejšie odhady nákladov a časového harmonogramu projektu, čím sa zabráni stavebným spoločnostiam špekulovať na základe neprehľadných údajov. To je výzva a zároveň príležitosť pre odvetvie, aby prehodnotilo svoju úlohu a prispôsobilo sa novému prostrediu, v ktorom sa transparentnosť a efektívnosť stanú kľúčovými faktormi úspechu.

Kľúčom k úspechu v ére digitálnej transformácie je ochota využiť vedomosti a uplatniť ich v praxi.

Spoločnosti, ktoré si to uvedomia ako prvé, budú mať v novej digitálnej konkurencii výhodu. Je však dôležité uvedomiť si, že samotné údaje nič nezmenia. Mnohí ľudia budú musieť zmeniť spôsob myslenia, a to si vyžaduje podnet. Vaša spoločnosť musí prehodnotiť spôsob zdieľania údajov

Ľudia, ktorí zmenia spoločnosť, sú tí, ktorí vedia s týmito údajmi pracovať, interpretovať ich, využiť ich na optimalizáciu a vytvoriť na ich základe novú architektúru procesov.

Ak čítate tieto riadky, ste pripravení na zmenu a už ste o krok vpred. Ďakujeme, že ste si vybrali túto cestu. Vitajte v ére digitálnej transformácie!



## O AUTOROVÍ

Volám sa Artyom Boiko. Moja cesta na stavenisko sa začala v roku 2007 - prácou baníka v bani na ťažbu bridlice v mojom rodnom meste, počas štúdia na Baníckej univerzite v Petrohrade v odbore banské a podzemné staviteľstvo. Na zadnej strane obálky tejto knihy môžete vidieť explorátora v čelbe, kde sme ťažili a odstrelovali stovky kubíkov ropnej bridlice. Moja kariéra sa uberala rôznymi smermi, od práce banského a podzemného staviteľa až po priemyselného horolezca, pokrývača a montéra výťahových zariadení. Mal som tú čest podieľať sa na projektoch rôzneho rozsahu, od výstavby súkromných domov až po veľké priemyselné zariadenia v rôznych regiónoch sveta.



Postupom času sa moja práca posunula od fyzickej výstavby k správe informácií a digitálnym procesom. Od roku 2013 som pracoval na rôznych pozíciách v malých, stredných a veľkých stavebných spoločnostiach vo viacerých regiónoch Nemecka, od projektanta až po manažéra správy dát. Pokial' ide o správu údajov, moje skúsenosti pozostávajú z práce s údajmi v rôznych ERP systémoch, CAD (BIM), MEP, FEM, CMS. Venoval som sa optimalizácii, automatizácii procesov, ako aj analýze, strojovému učeniu, spracovaniu údajov vo fázach plánovania, výpočtu a realizácie stavebných prác v priemyselných, bytových, infraštruktúrnych a komunálnych stavebných spoločnostiach.

So softvérom s otvoreným zdrojovým kódom a otvorenými údajmi pracujem od roku 2003. Počas tohto obdobia som realizoval mnoho webových projektov - od webových stránok a internetových obchodov až po plnohodnotné webové aplikácie - s využitím riešení s otvoreným zdrojovým kódom a otvorených CMS. Tieto platformy, v mnom podobné moderným stavebným ERP, majú modulárnu architektúru, vysokú prispôsobivosť a prístupnosť. Tieto skúsenosti položili základy môjho profesionálneho prístupu - zameranie na open source technológie a kultúru spoločného vývoja. Rešpektovanie open source a slobodná výmena poznatkov je niečo, čo sa snažím presadzovať v stavebníctve. Moja práca na zlepšení dostupnosti údajov v stavebníctve sa premietla do vytvorenia niekoľkých komunit na sociálnych sieťach, ktoré diskutujú o otvorenosti údajov a využívaní otvoreného softvéru v stavebníctve, ako aj do spustenia niekoľkých startupov vyvíjajúcich riešenia na zabezpečenie prístupu k údajom z rôznych uzavretých systémov a platforem.

Môj prínos pre odbornú komunitu spočíva v účasti ako prednášateľa na konferenciách venovaných interoperabilite CAD (BIM), ERP, 4D-5D, strojovému učeniu LLM a umelej inteligencii, ako aj v článkoch uverejnených v európskych publikáciách o stavebníctve. Jedným z mojich významných úspechov je vytvorenie "História BIM" [111], komplexnej mapy dôležitých softvérových riešení na správu údajov v stavebníctve. Môj 7-dielny seriál článkov "Vývoj BIM a lobistické hry", preložený do viacerých jazykov, bol všeobecne uznávaný ako pokus o objasnenie skrytej dynamiky digitálnych noriem.

Takto som prešiel od ťažby kameňa k ťažbe a systematizácii stavebných údajov. Som vždy otvorený odbornému dialógu, novým nápadom a spoločným projektom. S vďakou prijmem akúkoľvek spätnú väzbu a teším sa na vaše správy alebo na to, že vás uvidím medzi svojimi sledovateľmi na sociálnych sieťach. Ďakujem vám, že ste túto knihu dočítali až do konca! Budem rád, ak vám táto kniha pomôže lepšie pochopiť tému údajov v stavebníctve.

## **ĎALŠÍ VZŤAH**

Názory čitateľov zohrávajú dôležitú úlohu pri ďalšom vývoji publikácií a výbere prioritných témy. Obzvlášť cenné sú pripomienky k tomu, ktoré myšlienky sa ukázali ako užitočné a ktoré vyvolali pochybnosti a vyžadujú si ďalšie objasnenie alebo uvedenie zdrojov. Kniha obsahuje širokú škálu materiálov a analýz, z ktorých niektoré sa môžu zdať kontroverzné alebo subjektívne. Ak v priebehu čítania nájdete nepresnosti, nesprávne citované zdroje, logické nezrovnalosti alebo preklepy, budem vám vďačný za vaše pripomienky, myšlienky alebo kritiku, ktoré môžete poslať na adresu:[boikoartem@gmail.com](mailto:boikoartem@gmail.com). Alebo prostredníctvom správ na LinkedIn:[linkedin.com/in/boikoartem](https://linkedin.com/in/boikoartem).

Budem vďačný za každú zmienku o knihe Data-Driven Construction na sociálnych sieťach - zdieľanie skúseností s čítaním pomáha šíriť informácie o otvorených údajoch a nástrojoch a podporuje moju prácu.

## **KOMENTÁR K PREKLADU**

Táto kniha bola preložená pomocou technológie umelej inteligencie. Vďaka tomu sa proces prekladu výrazne zrýchli. Ako pri každej technologickej operácii sa však môžu vyskytnúť chyby alebo nepresnosti. Ak si všimnete niečo, čo sa vám zdá byť nesprávne alebo nesprávne preložené, napíšte mi, prosím, e-mail. Vaše pripomienky pomôžu zlepšiť kvalitu prekladu.

## **KOMUNITY V OBLASTI VÝSTAVBY ZALOŽENÉ NA ÚDAJOCH**

Je to miesto, kde môžete slobodne klásť otázky a zdieľať svoje problémy a riešenia:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: <https://www.youtube.com/@datadrivenconstruction>

## ĎALŠIE ZRUČNOSTI A POJMY

Okrem kľúčových princípov práce s údajmi v stavebnictve sa DataDrivenConstruction venuje aj širokej škále ďalších konceptov, programov a zručností, ktoré sú pre profesionálov pracujúcich s údajmi nevyhnutné. Niektoré z nich sú uvedené len v prehľade, ale pre prax sú kľúčové.

Čitateľ, ktorý má záujem, môže navštíviť webovú stránku DataDrivenConstruction.io, kde nájde odkazy na ďalšie materiály o kľúčových zručnostiach. Tieto materiály zahŕňajú prácu s jazykmi Python a Pandas, budovanie ETL -procesov, príklady spracovania údajov v stavebných CAD projektoch, systémy big data a moderné prístupy k vizualizácii a analýze stavebných údajov.

Pri príprave knihy "DataDrivenConstruction" a všetkých prípadových štúdií bolo použitých mnoho nástrojov a softvéru s otvoreným zdrojovým kódom. Autor by chcel podakovať vývojárom a spoluautorom týchto riešení:

- Python a Pandas - základ manipulácie s údajmi a automatizácie
- Scipy, NumPy, Matplotlib a Scikit-Learn - knižnice na analýzu dát a strojové učenie
- SQL a Apache Parquet - nástroje na ukladanie a spracovanie veľkého množstva stavebných údajov
- Otvorený zdroj CAD (BIM) nástroje na spracovanie otvorených údajov v otvorených formátoch
- N8n, Apache Airflow, Apache NiFi - systémy na orchestráciu a automatizáciu pracovných postupov
- DeepSeek, LlaMa, Mistral - Open Source LLM

Osobitné podakovanie patrí všetkým účastníkom diskusií na tému otvorených údajov a nástrojov v odborných komunitách a sociálnych sieťach, ktorých kritika, pripomienky a nápady pomohli zlepšiť obsah a štruktúru tejto knihy.

Sledujte vývoj projektu na webovej stránke DataDrivenConstruction.io, kde sú uverejňované nielen aktualizácie a opravy knihy, ale aj nové kapitoly, návody a praktické príklady použitia opísaných techník.

## MAXIMÁLNE POHODLIE S TLAČENOU VERZIOU

Držíte v rukách bezplatnú digitálnu verziu knihy **Data-Driven Construction**. Pre pohodlniešiu prácu a rýchly prístup k materiálom vám odporúčame venovať pozornosť **tlačenému vydaniu**:



■ **Vždy po ruke:** kniha v tlačenej podobe sa stane spoľahlivým pracovným nástrojom, ktorý vám umožní rýchlo nájsť a použiť správne vizualizácie a diagramy v akejkoľvek pracovnej situácii

■ **Vysoká kvalita ilustrácií:** všetky obrázky a grafika v tlačenom vydani sú prezentované v maximálnej kvalite

■ **Rýchly prístup k informáciám:** jednoduchá navigácia, možnosť robiť si poznámky, záložky a pracovať s knihou kdekoľvek.

Zakúpením plnej tlačenej verzie knihy získate praktický nástroj na pohodlnú a efektívnu prácu s informáciami: možnosť rýchlo využívať vizuálne materiály pri každodenných úlohách, rýchlo nájsť potrebné schémy a robiť si poznámky. Okrem toho svojou kúpou podporíte šírenie otvorených poznatkov.

Tlačenú verziu knihy si môžete objednať na adrese: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## JEDINEČNÁ PRÍLEŽITOSŤ PRE STRATEGICKÉ UMIESTNENIE

Ponúkame vám možnosť umiestniť reklamné materiály do bezplatnej verzie DataDrivenConstruction. Platená verzia publikácie už prvý rok po vydaní prilákala pozornosť odborníkov z viac ako 50 krajín sveta - od Latinskej Ameriky až po ázijsko-tichomorský región. Ak chcete prerokovať individuálne podmienky spolupráce a získať podrobne informácie o možnostiach umiestnenia, vyplňte, prosím, formulár späťnej väzby na oficiálnom portáli [datadrivenconstruction.io](http://datadrivenconstruction.io) alebo napíšte na kontakty uvedené na konci knihy.



## KAPITOLY KNIHY SÚ K DISPOZÍCII NA DATADRIVENCONSTRUCTION.IO

Kapitoly z knihy Data-Driven Construction si môžete prečítať na webovej stránke Data-Driven Construction, kde sú postupne zverejňované časti knihy, aby ste mohli rýchlo nájsť potrebné informácie a využiť ich pri svojej práci. Nájdete tu aj mnoho ďalších publikácií na podobné témy, ako aj príklady aplikácií a riešení, ktoré vám pomôžu rozvíjať vaše zručnosti a aplikovať údaje v stavebníctve.



## NAJNOVŠIE VERZIE KNIHY STIAHNUŤ Z OFICIÁLNEJ WEOVEJ STRÁNKY

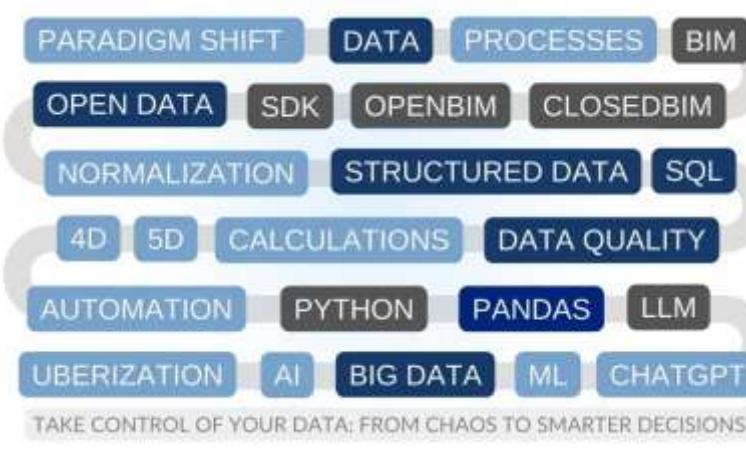
Aktuálne a najnovšie verzie knihy DataDrivenConstruction sú k dispozícii na stiahnutie na stránke [datadrivenconstruction.io](http://datadrivenconstruction.io). Ak chcete dostávať aktualizácie s novými kapitolami knihy, praktické tipy alebo recenzie nových aplikácií, prihláste sa na odber noviniek:

- Ako prví sa zoznámite s novými časťami knihy
- Získejte praktické prípadové štúdie a rady týkajúce sa analytiky a automatizácie v stavebníctve
- Aktuálne informácie o trendoch, publikáciách a príkladoch použitia

Prejdite na [datadrivenconstruction.io](http://datadrivenconstruction.io) a prihláste sa na odber!

# KONŠTRUKCIA ZALOŽENÁ NA ÚDAJOCH: PORADENSTVO, WORKSHOPY A ŠKOLENIA

Školiace programy a poradenstvo DataDrivenConstruction pomohli desiatkam popredných stavebných spoločností na celom svete zvýšiť efektivitu, znížiť náklady a zlepšiť kvalitu riešení. Medzi klientov DataDrivenConstruction patria niektorí z najväčších hráčov na miliardovom trhu vrátane stavebných, poradenských a IT spoločností.



Prečo si vybrať práve nás?

- **Relevantnosť:** rozprávanie o hlavných trendoch a poznatkoch v odvetví
- **Praktické:** pomáha profesionálom efektívne riešiť každodenné úlohy prostredníctvom PoC.
- **Individuálny prístup:** prispôsobený vašej firme, maximalizácia prínosov školení a poradenstva

Hlavné oblasti zamerania tímu DataDrivenConstruction:

- **Riadenie kvality údajov:** pomoc pri parametrizácii úloh, zhromažďovaní požiadaviek, overovanie a príprave údajov na automatizované spracovanie.
- **Data Mining - extrakcia a štruktúrovanie údajov:** nastavenie procesov ETL a extrakcia údajov z e-mailov, PDF, Excel, obrázkov a iných zdrojov.
- **Analytika BIM a CAD :** zhromažďovanie, štruktúrovanie a analýza informácií zo súborov RVT, IFC, DWG a iných formátov CAD (BIM).
- **Analytika a transformácia údajov:** premena rôznorodých informácií na štruktúrované údaje, analytiku, poznatky a rozhodnutia.
- **Integrácia údajov a automatizácia procesov:** od automatizovaného vytvárania dokumentov až po integráciu s internými systémami a externými databázami.

Kontaktujte DataDrivenConstruction.io a dozviete sa, ako môže automatizácia pomôcť vašej spoločnosti dosiahnuť hmatateľné obchodné výsledky.

## SLOVNÍČEK

**AI (Artificial Intelligence)** - umelá inteligencia; schopnosť počítačových systémov vykonávať úlohy, ktoré si bežne vyžadujú ľudskú inteligenciu, ako napríklad rozpoznávanie vzorov, učenie a rozhodovanie.

**Apache Airflow** je open source platforma na orchestráciu pracovných postupov, ktorá umožňuje programovo vytvárať, plánovať a sledovať pracovné postupy a ETL pomocou DAG (smerovaných acyklických grafov).

**Apache NiFi** je nástroj na automatizáciu dátových tokov medzi systémami, ktorý sa špecializuje na smerovanie a transformáciu údajov.

**Apache Parquet** je efektívny formát súborov na ukladanie stípcových údajov optimalizovaný na použitie v systémoch na analýzu veľkých objemov údajov. Poskytuje výraznú kompresiu a rýchle spracovanie.

**API (Application Programming Interface)** - formalizované rozhranie, ktoré umožňuje interakciu jedného programu s druhým bez prístupu k zdrojovému kódu, výmenu údajov a funkcií prostredníctvom štandardizovaných požiadaviek a odpovedí.

**Atribút** - charakteristika alebo vlastnosť objektu, ktorá opisuje jeho vlastnosti (napr. plocha, objem, cena, materiál).

**Databázy** sú organizované štruktúry na ukladanie, správu a prístup k informáciám, ktoré sa používajú na efektívne vyhľadávanie a spracovanie údajov.

**BEP (BIM Execution Plan)** - plán implementácie informačného modelovania budov, ktorý definuje ciele, metódy a procesy implementácie BIM v projekte.

**Veľké dáta (Big Data)** - súbory informácií s veľkým objemom, rôznorodosťou a rýchlosťou aktualizácie, ktoré si vyžadujú špeciálne technológie na spracovanie a analýzu.

**BI (Business Intelligence)** - Business **Intelligence**; procesy, technológie a nástroje na transformáciu údajov na zmysluplné informácie pre rozhodovanie.

**BIM (Building Information Modeling)** - informačné modelovanie budov; proces vytvárania a správy digitálnych reprezentácií fyzických a funkčných vlastností stavebných projektov, ktorý zahŕňa nielen 3D modely, ale aj informácie o vlastnostiach, materiáloch, čase a nákladoch.

**BlackBox/WhiteBox** - prístupy k pochopeniu systému: v prvom prípade je vnútorná logika skrytá, viditeľné sú len vstupy a výstupy; v druhom prípade je spracovanie transparentné a dostupné na analýzu.

**Bounding Box** je geometrická konštrukcia, ktorá opisuje hranice objektu v trojrozmernom priestore pomocou minimálnych a maximálnych súradníc X, Y a Z, čím vytvára "krabicu" okolo objektu.

**BREP (Boundary Representation)** je geometrická reprezentácia objektov, ktorá ich definuje prostredníctvom hraníc povrchov.

**CAD (Computer-Aided Design)** je systém počítačom podporovaného navrhovania, ktorý sa používa na vytváranie, úpravu a analýzu presných výkresov a 3D modelov v architektúre, stavebnictve, strojárstve a iných odvetviach.

**CAFM (Computer-Aided Facility Management)** je softvér na správu majetku a infraštruktúry, ktorý zahŕňa plánovanie priestorov, správu majetku, údržbu a monitorovanie nákladov.

**CDE (Common Data Environment)** - centralizovaný digitálny priestor na správu, ukladanie, zdieľanie a spoluprácu s informáciami o projekte vo všetkých fázach životného cyklu zariadenia.

**Centrum excelentnosti (CoE)** je špecializovaná štruktúra v rámci organizácie, ktorá je zodpovedná za rozvoj špecifickej oblasti znalostí, vývoj noriem a osvedčených postupov, školenie zamestnancov a podporu zavádzania inovácií.

**CoClass** je moderný systém klasifikácie stavebných prvkov tretej generácie.

**Konceptuálny dátový model** je vysokoúrovňové zobrazenie základných entít a ich vzťahov bez detailov atribútov, ktoré sa používa v počiatočných fázach návrhu databázy.

**CRM (Customer Relationship Management)** je systém riadenia vzťahov so zákazníkmi, ktorý sa používa na automatizáciu procesov predaja a služieb.

**DAG (Directed Acyclic Graph)** je smerový acyklický graf, ktorý sa používa v systémoch na orchestráciu dát (Airflow, NiFi) na určenie postupnosti úloh a závislostí.

**Dash** je rámcový Pythonový rámec na vytváranie interaktívnych webových vizualizácií údajov.

**Prístrojová doska - Prístrojová doska**, ktorá vizuálne prezentuje kľúčové ukazovatele výkonnosti a metriky v reálnom čase.

**Prístup zameraný na údaje** je metodika, ktorá uprednostňuje údaje pred aplikáciami alebo softvérovým kódom a robí z údajov ústredný majetok organizácie.

**Správa údajov** - súbor postupov, procesov a politík, ktoré zabezpečujú vhodné a efektívne využívanie údajov v rámci organizácie vrátane kontroly prístupu, kvality a bezpečnosti.

**Dátové jazero** je úložisko určené na ukladanie veľkého množstva nespracovaných údajov v pôvodnom formáte až do ich použitia.

**Data Lakehouse** je architektonický prístup, ktorý kombinuje flexibilitu a škálovateľnosť dátových jazier (**Data Lake**) so správou a výkonnosťou dátových skladov (DWH).

**Výstavba riadená údajmi** je strategický prístup, v ktorom je každá fáza životného cyklu stavby - od návrhu až po prevádzku - podporovaná automatizovanými, vzájomne prepojenými systémami. Tento prístup poskytuje nepretržité učenie sa na základe faktov, znižuje neistotu a umožňuje spoločnostiam dosiahnuť udržateľné vedúce postavenie v odvetví.

**Data-Driven Integrator** - spoločnosť, ktorá sa špecializuje na kombinovanie údajov z rôznych zdrojov a ich analýzu na účely manažérskych rozhodnutí.

**Prístup založený na údajoch** - metodika, pri ktorej sa údaje považujú za strategické aktívum a rozhodnutia sa prijímajú na základe objektívnej analýzy informácií, a nie subjektívnych názorov.

**Minimalizmus údajov** - prístup k redukcii údajov na najhodnotnejšie a najvýznamnejšie, ktorý umožňuje zjednodušené spracovanie a analýzu informácií.

**Dátový močiar** - rozptýlené množstvo neštruktúrovaných údajov, ktoré vzniká, keď sa informácie zhromažďujú a ukladajú nekontrolovaným spôsobom bez náležitej organizácie.

**DataOps** je metodika, ktorá spája princípy DevOps, dát a analytiku a zameriava sa na zlepšenie spolupráce, integrácie a automatizácie dátových tokov.

**Digitalizácia informácií** je proces prevodu všetkých aspektov stavebnej činnosti do digitálnej podoby vhodnej na analýzu, interpretáciu a automatizáciu.

**DataFrame** - Dvojrozmerná tabuľková dátová štruktúra v knižnici Pandas, kde riadky predstavujú jednotlivé záznamy alebo objekty a stĺpce ich vlastnosti alebo atribúty.

**Deskriptívna analýza** - Analýza historických údajov s cieľom pochopiť, čo sa stalo v minulosti.

**Diagnostická analýza** - Analýza údajov s cieľom zistiť, prečo sa niečo stalo.

**Ganttov diagram** je nástroj na plánovanie projektu, ktorý predstavuje úlohy ako vodorovné stĺpce na časovej osi a umožňuje vám vizualizovať postupnosť a trvanie práce.

**DWH (Data Warehouse)** je centralizovaný systém dátového skladu, ktorý zhromažďuje informácie z viacerých zdrojov, štruktúruje ich a sprístupňuje na účely analýzy a reportovania.

**ESG (Environmental, Social, Governance)** - súbor kritérií na posudzovanie environmentálnych, sociálnych a riadiacich vplyvov spoločnosti alebo projektu.

**ELT (Extract, Load, Transform)** je proces, pri ktorom sa údaje najprv extrahujú zo zdrojov, načítajú do úložiska a potom sa transformujú na analytické účely.

**ETL (Extract, Transform, Load)** je proces extrakcie údajov z rôznych zdrojov, ich transformácie do požadovaného formátu a načítania do cieľového úložiska na analýzu.

**ER-diagram (Entity-Relationship)** - vizuálny diagram zobrazujúci entity, ich atribúty a vzťahy medzi nimi, ktorý sa používa pri modelovaní údajov.

**ERP (Enterprise Resource Planning)** je komplexný modulárny systém plánovania podnikových zdrojov, ktorý sa používa na riadenie a optimalizáciu rôznych aspektov stavebného procesu.

**Funkcie** - v strojovom učení nezávislé premenné alebo atribúty používané ako vstupy do modelu.

**Fyzický dátový model** - podrobnejšia reprezentácia štruktúry databázy vrátane tabuľiek, stĺpcov, dátových typov, klúčov a indexov, optimalizovaná pre konkrétny DBMS.

**FPDF** je knižnica jazyka Python na vytváranie dokumentov PDF.

**Geometrické jadro** je softvérový komponent, ktorý poskytuje základné algoritmy na vytváranie, úpravu a analýzu geometrických objektov v CAD, BIM a iných inžinierskych aplikáciách.

**HiPPO (Highest Paid Person's Opinion)** - prístup k rozhodovaniu založený skôr na názore najlepšie platenej osoby v organizácii než na objektívnych údajoch.

**IDE (Integrated Development Environment)** - integrované vývojové prostredie, komplexný nástroj na písanie, testovanie a ladenie kódu (napr. PyCharm, VS Code, Jupyter Notebook).

**IDS (Information Delivery Specification)** je špecifikácia poskytovania informácií, ktorá definuje požiadavky na údaje v rôznych fázach projektu.

**IFC (Industry Foundation Classes)** je formát na výmenu údajov BIM, ktorý zabezpečuje interoperabilitu medzi rôznymi softvérovými riešeniami.

**Priemysel 5.0** je koncepcia rozvoja priemyslu, ktorá spája možnosti digitalizácie, automatizácie a umelej inteligencie s ľudským potenciálom a environmentálnou udržateľnosťou.

**Integrácia údajov** je proces spájania údajov z rôznych zdrojov do jedného uceleného systému, ktorý poskytuje jednotný pohľad na informácie.

**Informačné silá** sú izolované systémy na ukladanie údajov, ktoré nezdieľajú informácie s inými systémami, čo vytvára prekážky pre efektívne využívanie údajov.

**Internet vecí (Internet of Things)** je koncept pripojenia fyzických objektov k internetu s cieľom zhromažďovať, spracovávať a prenášať údaje.

**k-NN (k-Nearest Neighbors)** je algoritmus strojového učenia, ktorý klasifikuje objekty na základe podobnosti s najbližšími susedmi v trénovacej vzorke.

**Kaggle** je platforma na analýzu údajov a súťaže v strojovom učení.

**Kalkulácia** - výpočet nákladov na stavebné práce alebo procesy pre určitú mernú jednotku (napr. 1 m<sup>2</sup> sadrokartónovej steny, 1 m<sup>3</sup> betónu).

**KPI (Key Performance Indicators)** sú **kľúčové** ukazovatele výkonnosti, merateľné ukazovatele, ktoré sa používajú na hodnotenie úspešnosti spoločnosti alebo konkrétneho projektu.

**Štítky** - v strojovom učení cielové premenné alebo atribúty, ktoré by mal model predpovedať.

**Algoritmus učenia** - proces hľadania najlepšej hypotézy v modeli zodpovedajúcej cielovej funkcií pomocou súboru trénovacích údajov.

**Lineárna regresia** - štatistická metóda modelovania vzťahu medzi závislou premennou a jednou alebo viacerými nezávislými premennými.

**LLM (Large Language Model)** - Veľký jazykový model, umelá inteligencia vycvičená na pochopenie a generovanie textu z obrovského množstva údajov, schopná analyzovať kontext a písat programový kód.

**LOD (Level of Detail/Development)** - úroveň podrobnosti modelu, ktorá určuje stupeň geometrickej presnosti a informačného obsahu.

**Logický dátový model** je podrobny opis entít, atribútov, kľúčov a vzťahov, ktorý odráža obchodné informácie a pravidlá a je medzistupňom medzi konceptuálnym a fyzickým modelom.

**Strojové učenie** - trieda techník umelej inteligencie, ktoré umožňujú počítačovým systémom učiť sa a predpovedať na základe údajov bez explicitného programovania.

**Masterformat** je klasifikačný systém prej generácie, ktorý sa používa na štruktúrovanie stavebných špecifikácií do sekcií a disciplín.

**MEP (Mechanical, Electrical, Plumbing)** - technické systémy budov, ktoré zahŕňajú mechanické, elektrické a vodovodné komponenty.

**Siet** je reprezentácia 3D objektov pozostávajúca z vrcholov, hrán a plôch.

**Model** - V strojovom učení súbor rôznych hypotéz, z ktorých jedna approximuje cielovú funkciu, ktorá sa má predpovedať alebo approximovať.

**Modelovanie údajov** je proces vytvárania štruktúrovanej reprezentácie údajov a ich vzťahov na implementáciu v informačných systémoch vrátane konceptuálnej, logickej a fyzickej úrovne.

**n8n** je open source nástroj na automatizáciu pracovných postupov a integráciu aplikácií prostredníctvom nízkokódového prístupu.

**Normalizácia** - v strojovom učení proces zosúladenia rôznych číselných údajov do spoločnej mierky s cieľom uľahčiť ich spracovanie a analýzu.

**Reverzné inžinierstvo** - proces skúmania zariadenia, fungovania a výrobnej technológie objektu prostredníctvom analýzy jeho štruktúry, funkcií a fungovania. V kontexte údajov - extrakcia informácií z proprietárnych formátov na použitie v otvorených systémoch.

**OCR (Optical Character Recognition)** je technológia **optického** rozpoznávania **znakov**, ktorá konvertuje textové obrázky (naskenované dokumenty, fotografie) do strojovo čitateľného textového formátu.

**OmniClass** je medzinárodný klasifikačný štandard druhej generácie pre správu informácií o stavbách.

**Ontológia** - systém vzájomných vzťahov pojmov, ktorý formalizuje určitú oblast poznania.

**Open Source** - model vývoja a distribúcie softvéru s otvoreným zdrojovým kódom, ktorý je k dispozícii na bezplatné používanie, štúdium a modifikáciu.

**Open BIM** je koncept otvoreného BIM, ktorý zahŕňa používanie otvorených štandardov a formátov na výmenu údajov medzi rôznymi softvérovými riešeniami.

**Otvorené štandardy** - verejne dostupné špecifikácie na dosiahnutie konkrétnej úlohy, ktoré umožňujú rôznym systémom vzájomnú interakciu a výmenu údajov.

**Pandas** je open source knižnica jazyka Python na spracovanie a analýzu údajov, ktorá poskytuje dátové štruktúry DataFrame a Series na efektívne spracovanie tabuľkových informácií.

**Paradigma otvorených údajov** je prístup k spracovaniu údajov, pri ktorom sú informácie voľne dostupné na použitie, opäťovné použitie a šírenie každému.

**Parametrická** metóda je metóda odhadu stavebného projektu, ktorá využíva štatistické modely na odhad hodnoty na základe parametrov projektu.

**PIMS (Project Information Model)** je digitálny systém určený na organizáciu, ukladanie a zdieľanie všetkých informácií o projekte.

**Pipeline** - postupnosť procesov spracovania údajov od extrakcie a transformácie až po analýzu a vizualizáciu.

**PMIS (Project Information Management System)** je systém riadenia projektov určený na podrobné riadenie úloh na úrovni jednotlivých stavebných projektov.

**Prediktívna** analýza je časť analytiky, ktorá využíva štatistické metódy a strojové učenie na predpovedanie budúcich výsledkov na základe historických údajov.

**Preskriptívna** analýza - časť analytiky, ktorá nielen predpovedá budúce výsledky, ale aj navrhuje optimálne kroky na dosiahnutie požadovaných výsledkov.

**Proprietárne** formáty sú uzavreté dátové formáty kontrolované konkrétnou spoločnosťou, ktoré obmedzujú možnosť zdieľania informácií a zvyšujú závislosť od konkrétneho softvéru.

**QTO (Quantity Take-Off)** je proces získavania množstiev prvkov z projektovej dokumentácie s cieľom vypočítať množstvá materiálov potrebných na realizáciu projektu.

**Systém manažérstva kvality** - systém manažérstva kvality, ktorý zabezpečuje, že procesy a výsledky splňajú stanovené požiadavky.

**RAG (Retrieval-Augmented Generation)** je metóda, ktorá kombinuje generatívne schopnosti jazykových modelov s extrakciou relevantných informácií z podnikových databáz, čím zlepšuje presnosť a relevantnosť odpovedí.

**RDBMS (Relational Database Management System)** je relačný systém správy databáz, ktorý organizuje informácie vo forme vzájomne prepojených tabuľiek.

**RegEx (regulárne výrazy)** je formalizovaný jazyk na vyhľadávanie a spracovanie reťazcov, ktorý umožňuje špecifikovať šablóny na kontrolu textových údajov na splnenie určitých kritérií.

**Regresia** je štatistická metóda analýzy vzťahu medzi premennými.

**Výpočty CO<sub>2</sub>** sú metódou odhadu emisií oxidu uhličitého spojených s výrobou a používaním stavebných materiálov a procesov.

**Metóda zdrojov** je metóda odhadu založená na podrobnej analýze všetkých zdrojov (materiálov, práce, zariadení) potrebných na vykonanie stavebných prác.

**RFID (Radio Frequency Identification)** je technológia na automatickú identifikáciu objektov pomocou rádiových signálov, ktorá sa používa na sledovanie materiálov, strojov a zamestnancov.

**ROI (Return on Investment)** je ukazovateľ vyjadrujúci pomer medzi ziskom a investovanými prostriedkami, ktorý sa používa na hodnotenie efektívnosti investícií.

**SaaS (softvér ako služba)** je model softvéru ako služby, pri ktorom sú aplikácie hostované poskytovateľom a sprístupnené používateľom prostredníctvom internetu.

**SCM (Supply Chain Management)** - riadenie dodávateľského reťazca, ktoré zahŕňa koordináciu a optimalizáciu všetkých procesov od obstarávania materiálov až po dodávku hotových výrobkov.

**Dátové silá** sú izolované úložiská informácií v rámci organizácie, ktoré nie sú integrované s inými systémami, čo sťaže zdieľanie údajov a je neefektívne.

**SQL** (Structured Query **Language**) je štruktúrovaný dopytovací jazyk používaný na prácu s relačnými databázami.

**SQLite** je ľahký, zabudovateľný, multiplatformový DBMS, ktorý nevyžaduje samostatný server a podporuje základné funkcie SQL, široko používaný v mobilných aplikáciach a zabudovaných systémoch.

**Štruktúrované údaje** - informácie usporiadane v určitom formáte s jasou štruktúrou, napríklad v relačných databázach alebo tabuľkách.

**Voľne štruktúrované údaje** - informácie s čiastočnou organizáciou a flexibilnou štruktúrou, napríklad JSON alebo XML, kde rôzne prvky môžu obsahovať rôzne sady atribútov.

Entita je konkrétny alebo abstraktný objekt reálneho sveta, ktorý možno jednoznačne identifikovať, opísat a reprezentovať vo forme údajov.

**Supervised Learning** - typ strojového učenia, pri ktorom sa algoritmus trénuje na rozdelených údajoch, príčom pre každý príklad je známy požadovaný výsledok.

**Taxonómia** je hierarchický klasifikačný systém, ktorý sa používa na systematickú kategorizáciu prvkov na základe spoločných znakov.

**Súbor údajov Titanic** je populárny súbor údajov na trénovanie a testovanie modelov strojového učenia.

**Trénovanie** - proces, v ktorom algoritmus strojového učenia analyzuje údaje s cieľom identifikovať vzory a vytvoriť model.

**Transferové učenie** je technika strojového učenia, pri ktorej sa model vycvičený pre jednu úlohu používa ako východiskový bod pre inú úlohu.

**Transformácia** - proces zmeny formátu, štruktúry alebo obsahu údajov na neskoršie použitie.

**Požiadavky na údaje** - formalizované kritériá definujúce štruktúru, formát, úplnosť a kvalitu informácií potrebných na podporu obchodných procesov.

**Uberizácia stavebníctva** je proces transformácie tradičných obchodných modelov v stavebníctve pod vplyvom digitálnych platform, ktoré poskytujú priamu interakciu medzi zákazníkmi a dodávateľmi bez sprostredkovateľov.

**Uniclass** je systém klasifikácie stavebných prvkov druhej a tretej generácie, ktorý sa široko používa v Spojenom kráľovstve.

**USD (Universal Scene Description)** je dátový formát vyvinutý pre počítačovú grafiku, ale vďaka svojej jednoduchej štruktúre a nezávislosti na geometrických jadrách sa uplatnil aj v technických systémoch.

**Validácia údajov** je proces kontroly informácií na základe stanovených kritérií a požiadaviek s cieľom zabezpečiť presnosť, úplnosť a konzistentnosť údajov.

**Vektorová databáza** - špecializovaný typ databázy, ktorá ukladá údaje ako viacozmerné vektory na efektívne sémantické vyhľadávanie a porovnávanie objektov.

**Vektorová reprezentácia (embedding)** je metóda transformácie údajov na viacozmerné číselné vektory, ktorá umožňuje strojovým algoritmom efektívne spracúvať a analyzovať informácie.

**VectorOps** je metodika zameraná na spracovanie, ukladanie a analýzu viacozmerných vektorových údajov, ktorá je dôležitá najmä v oblastiach, ako sú digitálne dvojčatá a sémantické vyhľadávanie.

**Vizualizácia** - grafické znázornenie údajov na lepšie vnímanie a analýzu informácií.

Abecedná kategorizácia pojmov sa uskutočnila podľa ich anglických názvov.



## ZOZNAM ODKAZOV A ONLINE MATERIÁLOV

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics", 12. decembra 2016. [Online]. Dostupné na: <https://www.gartner.com/en/documents/3530919>. [Dátum prístupu: 1. marec 2025].
- [2] KPMG, "Známe výzvy - nové prístupy. Globálny prieskum stavebníctva 2023," 1. január 2023. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Dostupné: 5. marca 2025].
- [3] F. R. Barnard, "Obraz je hodný tisícov slov", 10 Mapi 1927. [Online]. Dostupné na: [https://en.wikipedia.org/wiki/A\\_picture\\_is\\_worth\\_a\\_thousand\\_words](https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words). [Dátum odkazu: 15. marec 2025].
- [4] M. Bastian, "Generálny riaditeľ Microsoftu Satya Nadella hovorí, že samozvané AGI je "nezmyselný benchmark hacking"," 21. februára 2025. [Online]. Dostupné na: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Dátum prístupu: 15. marca 2025].
- [5] W. E. Fórum, "Fórum formuje budúcnosť stavebníctva - krajina v premene:", 1. januára 2016. [Online]. Dostupné na: [https://www3.weforum.org/docs/WEF\\_Shaping\\_the\\_Future\\_of\\_Construction.pdf](https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf). [Dátum prístupu: 2. marca 2025].
- [6] C. D. Gillespie, "Clay: Earth's entanglement in the age of clay", 2024. [Online]. Dostupné na: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Papyrus, 3. storočie pred n. l. Jazyk je grécky", 2024. [Online]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Monitorovanie: využitie dostupných nástrojov", 1980. [Online]. Dostupné na: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Dátum vystúpenia: 15. marec 2025].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world", 15. februára 2015. [Online]. Dostupné na: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Dátum prístupu: 15. marca 2025].
- [10] Skanska USA, "Trendy na jesennom stavebnom trhu", 2. novembra 2023. [Online]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Oxford Essential Quotations (4. vydanie)", Oxford University Press, 2016. [Online]. Available:

- <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-or-ed4-00006236>. [Dátum prístupu: 1. marec 2025].
- [12] "Citát: Sondergaard o analýze údajov," [Online]. Dostupné na: <https://www.causeweb.org/cause/resources/library/r2493>. [Dátum prístupu: 15. marec 2025].
- [13] "Ako globálny záujem o umelú inteligenciu zvyšuje trh so správou údajov", 28. mája 2024. [Online]. Dostupné na: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Dátum prístupu: 15. marca 2025].
- [14] И. McCue, "História ERP", 2024. [Online]. Dostupné na: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "Cena ERP: Kol'ko stojí systém ERP?," [Online]. Dostupné na: <https://erpscout.de/en/erp-costs/>. [Dátum prístupu: 15. marec 2025].
- [16] softwarepath, "Čo nám hovorí 1 384 projektov ERP o výbere ERP (správa ERP 2022)", 18. 1. 2022. [Online]. Dostupné na: <https://softwarepath.com/guides/erp-report>. [Dátum prístupu: 15. marca 2025].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16. decembra 2016. [Online]. Dostupné na: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Dátum oslovenia: 1. marec 2025].
- [18] McKinsey, "The data-driven enterprise of 2025", 28. januára 2022. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Dátum referencie: 22. mája 2024].
- [19] Wikipedia, "Moorov zákon," [Online]. Dostupné na: [https://en.wikipedia.org/wiki/Moore%27s\\_law](https://en.wikipedia.org/wiki/Moore%27s_law). [Dátum prístupu: 15. marec 2025].
- [20] Accenture, "Building More Value With Capital Projects", 1. januára 2020. [Online]. Dostupné na: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accenture-industryx-building-value-capital-projects-highres.pdf>. [Dátum prístupu: 3. marca 2024].
- [21] B. Marr, "Koľko údajov denne vytvoríme? Ohromujúce štatistiky, ktoré by si mal prečítať každý," 2018. [Online]. Dostupné na: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Koľko údajov sa denne vyprodukuje?", 2024. [Online]. Dostupné na: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.

- [23] T. Sullivan, "AI and the global 'datasphere': how much information will humanity have by 2025?" (Umelá inteligencia a globálna "dátová sféra": koľko informácií bude mať ľudstvo do roku 2025?), 2024. [Online]. Dostupné na: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Celkový počet tlačených kníh vydaných v rôznych regiónoch západnej Európy v každom polstoročí v rokoch 1454 až 1800," [internet]. [Online]. Dostupné na: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Dátum odkazu: 1. marec 2025].
- [25] "Cenové príklady", 2024. [Online]. Dostupné na: <https://cloud.google.com/storage/pricing-examples>.
- [26] M. Ashare, "Enterprises outsourcing data storage as complexity rises", 10. mája 2024. [Online]. Dostupné na: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Dátum prístupu: 15. marca 2025].
- [27] JETSOFTPRO, "SaaS je mŕtvy? Vysvetlenie šokujúcej predpovede generálneho riaditeľa spoločnosti Microsoft," 13. januára 2025. [Online]. Dostupné na: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Satya Nadella | BG2 w/ Bill Gurley & Brad Gerstner," 12. decembra 2024. [Online]. Dostupné na: [https://www.youtube.com/watch?v=9NtsnzRFJ\\_o](https://www.youtube.com/watch?v=9NtsnzRFJ_o). [Dátum prístupu: 15. marca 2025].
- [29] GoodReads, "Tim Berners-Lee," [online]. Dostupné na: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Dátum prístupu: 15. marec 2025].
- [30] KPMG, "Cue Construction 4.0: Make-or-Break Time", 1. januára 2023. [Online]. Dostupné na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Dostupné: 5. marca 2025].
- [31] I. Deininger, B. Koch, R. Bauknecht a M. Langhans, "Using digital models for decarbonising a production site: An example of connecting a building model, a production model and an energy model," 2024. [Online]. Dostupné na: [https://www.researchgate.net/publication/374023998\\_Using\\_Digital\\_Models\\_to\\_Decarbonize\\_a\\_Production\\_Site\\_A\\_Case\\_Stud...](https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Stud...)
- [32] McKinsey, "REINVENCIÁCIA STAVEBNÍCTVA: CESTA K VYŠŠEJ PRODUKTIVITE", 1. februára 2017. [Online]. Dostupné na: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insi>

- ghts/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf.
- [33] Pracovná skupina pre stavebníctvo podpredsedovi vlády, "Rethinking Construction", 1. októbra 2014. [Online]. Dostupné na: [https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking\\_construction\\_report.pdf](https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf).
- [34] Forbes, "Bez názoru ste len ďalším človekom s údajmi", 15. marca 2016. [Online]. Dostupné na: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Dátum prístupu: 15. marca 2025].
- [35] Wikiquote, "Charles Babbage," [online]. Dostupné na: [https://en.wikiquote.org/wiki/Charles\\_Babbage](https://en.wikiquote.org/wiki/Charles_Babbage). [Dátum prístupu: 15. marec 2025].
- [36] SAP, "Nový výskum zistil, že takmer polovica vedúcich pracovníkov dôveruje umelej inteligencii viac ako sebe", 12. marca 2025. [Online]. Dostupné na: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Dátum prístupu: 15. marca 2025].
- [37] Kanadská asociácia stavebníctva a KPMG v Kanade, 2021, "Stavebníctvo v digitálnom svete", 1. máj 2021. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Dátum prístupu: 5. marca 2025].
- [38] ZCS, "Dekódovanie piatej priemyselnej revolúcie," [Online]. Dostupné na: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Dátum prístupu: 15. marec 2025].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012.
- [40] F. N. a Y. Z. Harvard Business School: Manuel Hoffmann, "The Value of Open Source Software", 24. 1. 2024. [Online]. Dostupné na: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Dátum prístupu: 15. marca 2025].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Príručka na odhadovanie nákladov na vývoj softvéru", 1. septembra 2008. [Online]. Dostupné na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%20I%20rev%2010.pdf>.
- [42] McKinsey, "Zvyšovanie produktivity v stavebníctve" [online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Dátum oslovenia: 15. marec 2025].
- [43] A. G. a. C. Syverson, "The Strange and Awful Path of Productivity in the US Construction Sector",

19. 1. 2023. [Online]. Dostupné na: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Dátum prístupu: 1. marec 2025].
- [44] McKinsey, "Dodržiavanie produktivity v stavebníctve už nie je voliteľné", 9. augusta 2024. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Dátum prístupu: 5. marca 2025].
- [45] ING Group, "Zaostávanie produktivity v stavebníctve zvyšuje náklady na výstavbu", 12. decembra 2022. [Online]. Dostupné na: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Dátum oslovenia: 15. marca 2025].
- [46] M. Berman, "Šokujúca predpoveď generálneho riaditeľa spoločnosti Microsoft: "Agenti nahradia všetok softvér", 19. decembra 2024. [Online]. Dostupné na: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Dátum prístupu: 15. marca 2025].
- [47] Business Insider, "Generálny riaditeľ spoločnosti Anthropic tvrdí, že do 3 až 6 mesiacov bude umelá inteligencia písat' 90 % kódu, ktorý malí na starosti vývojári softvéru", 15. marca 2025. [Online]. Dostupné na: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Dátum prístupu: 30. marca 2025].
- [48] Statista, "Porovnanie popularity systémov na správu databáz (DBMS) na celom svete k júnu 2024, podľa kategórií", jún 2024. [Online]. Dostupné na: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Dátum oslovenia: 15. marca 2025].
- [49] DB-Engines, "DB-Engines Ranking," [online]. Dostupné na: <https://db-engines.com/en/ranking>. [Dátum odkazu: 15. marec 2025].
- [50] "Prieskum Stack Overflow medzi vývojármí 2023", 2024. [Online]. Dostupné na: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024. [Online]. Dostupné na: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Štruktúrované a neštruktúrované údaje: aký je medzi nimi rozdiel?", 2024. [Online]. Dostupné na: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF POROVNANIE FORMÁTOV ÚDAJOV PRE STAVEBNÉ PROJEKTY", 23. 4. 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Building Information Modeling Whitepaper site," 2003. [Online]. Dostupné na:

- <https://web.archive.org/web/20030711125527/http:/usa.autodesk.com/adsk/servlet/item?id=2255342&siteID=123112>.
- [55] A. Bojko, "Lobistické vojny a rozvoj BIM. Časť 5: BlackRock je pánom všetkých technológií. Ako korporácie ovládajú otvorený zdrojový kód," 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] D. Ushakov, "Priame modelovanie - kto a prečo ho potrebuje? Prehľad konkurenčných technológií", 14. 11. 2011. [Online]. Dostupné na: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Dátum oslovenia: 02 2025].
- [57] C. Eastman a A. Cthers, "Eastman, Charles; And Cthers", september 1974. [Online]. Dostupné na: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Dátum prístupu: 15. marec 2025].
- [58] D. Ushakov, "Priame modelovanie - kto a prečo ho potrebuje? Prehľad konkurenčných technológií", 11. novembra 2011. [Online]. Dostupné na: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Dátum odkazu: 15. marca 2025].
- [59] D. Weisberg, "História CAD", 12. decembra 2022. [Online]. Dostupné na: [https://www.shapr3d.com/blog/history-of-cad?utm\\_campaign=cadhISTORYnet](https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhISTORYnet). [Dátum odkazu: 15. marca 2025].
- [60] ADSK, "Biela kniha Informačné modelovanie budov", 2002. [Online]. Dostupné na: [https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http:/images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Dátum odkazu: 15. marec 2025].
- [61] ADSK, "Biela kniha Informačné modelovanie budov v praxi," [online]. Dostupné na: [https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac\\_sapac\\_main/files/4525077\\_BIM\\_in\\_Practice.pdf](https://web.archive.org/web/20060512181000/http:/images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf). [Dátum prístupu: 15. marec 2025].
- [62] A. Bojko, "Lobistické vojny a rozvoj BIM. Časť 2: otvorený BIM VS uzavretý BIM. Európa VS zvyšok sveta," 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs," 2024. [Online]. Dostupné na: [https://youtu.be/S-TNdUgfHxk?si=evM\\_v28KQbGOG0k&t=1360](https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360).
- [64] ADSK, "Biela kniha BIM", 2002. [Online]. Dostupné na: [https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http:/images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Dátum oslovenia: 15. marec 2025].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale," [online]. Dostupné

- na:  
[https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Dátum prístupu: 15. marec 2025].
- [66] M. Shacklett, "Štruktúrované a neštruktúrované údaje: Kľúčové rozdiely," 2024. [Online]. Dostupné na: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] K. Woolard, "Making sense of the growth of unstructured data", 2024. [Online]. Dostupné na: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] A. C. O. J. L. D. J. a. L. T. G. Michael P. Gallaher, "Cost Analysis of Inadequate Interoperability in the," 2004. [Online]. Dostupné na: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf> . [Dátum prístupu: 02 2025].
- [69] CrowdFlower, "Data Science Report 2016", 2016. [Online]. Dostupné na: [https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf). [Dátum prístupu: 15. marca 2025].
- [70] Analyticsindiamag, "6 časovo najnáročnejších úloh pre dátových vedcov", 15. mája 2019. [Online]. Dostupné na: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Správa: Dátoví vedci trávia väčšinu času upratovaním", 6. júla 2015. [Online]. Dostupné na: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Dátum prístupu: 5. marca 2025].
- [72] S. Hawking, "Science AMA Series: Stephen Hawking AMA Answers!", 27. júla 2015. [Online]. Dostupné na: [https://www.reddit.com/r/science/comments/3ny5i/science\\_ama\\_series\\_stephen\\_hawking\\_a\\_ma\\_answers/](https://www.reddit.com/r/science/comments/3ny5i/science_ama_series_stephen_hawking_a_ma_answers/). [Dátum prístupu: 15. marca 2025].
- [73] B. Cyphers a K. Doctorow, "Privacy without Monopoly: Data Protection and Interoperability", 2024. [Online]. Dostupné na: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", 1. októbra 2013. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Dátum prístupu: 15. marca 2025].
- [75] A. Bojko, "Boj o otvorené údaje v stavebníctve. História systémov AUTOLISP, intelliCAD, openDWG, ODA a openCASCADE," 15 05 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Dátum oslovenia: 16. 02. 2025].

- [76] Wikipédia, "Microsoft a open source," [online]. Dostupné na: [https://en.wikipedia.org/wiki/Microsoft\\_and\\_open\\_source](https://en.wikipedia.org/wiki/Microsoft_and_open_source). [Dátum prístupu: 15. marec 2025].
- [77] TIME, "The Gap Between Open and Closed AI Models Might Be Shrinking. Tu je dôvod, prečo je to dôležité," 5. novembra 2024. [Online]. Dostupné na: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Dátum prístupu: 15. marca 2025].
- [78] The Verge, "Viac ako štvrtina nového kódu v spoločnosti Google je generovaná umelou inteligenciou", 29. októbra 2024. [Online]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Dátum prístupu: 15. marca 2025].
- [79] McKinsey Digital, "Obchodné argumenty pre využitie GPU na zrýchlenie analytického spracovania", 15. decembra 2020. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-spracovanie>. [Dostupné: 15. marca 2025].
- [80] PwC, "PwC Open Source Monitor 2019", 2019. [Online]. Dostupné na: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Dostupné 15. marca 2025].
- [81] Travers Smith, "The Open Secret: Open Source Software", 2024. [Online]. Dostupné na: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Dátum prístupu: 15. marca 2025].
- [82] Deloitte, "Proces prenosu údajov pri transformácii podnikov", 2021. [Online]. Dostupné na: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Dátum referencie: 15. marec 2025].
- [83] gov.uk, "Analýza údajov a umelá inteligencia pri realizácii vládnych projektov", 20. marca 2024. [Online]. Dostupné na: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Dátum prístupu: 5. marca 2025].
- [84] "Citát Pôvod: Všetko by malo byť čo najjednoduchšie, ale nie jednoduchšie," 13. mája 2011. [Online]. Dostupné na: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Dátum prístupu: 15. marca 2025].
- [85] "Transformátor (architektúra hlbokého učenia)," [Online]. Dostupné na: [https://en.wikipedia.org/wiki/Transformer\\_\(deep\\_learning\\_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Dátum prístupu: 15. marca 2025].
- [86] "Štatistiky sťahovania balíkov Python", 2024. [Online]. Dostupné na: <https://www.pepy.tech/projects/pandas>.

- [87] Rozhovor Bit, "Top 10 knižníc Pythonu", 2023. [Online]. Dostupné na: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%2013%2C000%20libraries,data%20manipulation%2C%20and%20many%20more>. [Dátum prístupu: 30. marca 2025].
- [88] "NVIDIA a HP podporujú dátovú vedu a generatívnu AI na pracovných staniciach," 7. marca 2025. [Online]. Dostupné na: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Dátum prístupu: 15. marca 2025].
- [89] P. Orac, "Ako spracovať DataFrame s miliónmi riadkov za niekoľko sekúnd", 2024. [Online]. Dostupné na: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Uslu, "Čo je Kaggle?", 2024. [Online]. Dostupné na: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "Prednáška generálneho riaditeľa spoločnosti NVIDIA Jensena Huangu na veľtrhu COMPUTEX 2024", 2. júna 2024. [Online]. Dostupné na: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Dostupné 15. marca 2025].
- [92] "Členovia: zakladatelia a firemní členovia", 2024. [Online]. Dostupné na: <https://www.opendesign.com/member-showcase>.
- [93] A. Bojko, "Vek zmien: IFC je minulosťou alebo prečo sa ADSK a ďalší dodávateľia CAD chcú vzdať IFC za USD v 14 kľúčových faktoch", 24. novembra 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82ccd10a>. [Dostupné: 23. februára 2025].
- [94] A. Bojko, "Svet po BIM. Prechod na dáta a procesy a či stavebnictvo potrebuje sémantiku, formáty a interoperabilitu," 20. 12. 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Dátum vystúpenia: 23. februára 2025].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016. [Online]. Dostupné na: [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). [Dátum prístupu: 23. februára 2025].
- [96] Harvard Business Review, "Zlé údaje stoja USA 3 bilióny dolárov ročne", 22. septembra 2016. [Online]. Dostupné na: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Dopady kvality údajov", 1. januára 2025. [Online]. Dostupné na: <https://delpha.io/impacts-of-data-quality/>.

- [98] W. B. D. Guide, "Design for Maintainability: The Importance of Operations and Maintenance Considerations During the Design Phase of Construction Projects" (Projektovanie pre udržiavateľnosť: Význam úvah o prevádzke a údržbe vo fáze projektovania stavebných projektov)," [Online]. Dostupné na: <https://www.wbdg.org/resources/design-for-maintainability>. [Dátum prístupu: 15. marec 2025].
- [99] O. o. D. C. P. a. Oversight, "Príručka pre plánovanie prevencie a kontroly korózie pre vojenské systémy a zariadenia", apríl 2014. [Online]. Dostupné na: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guide%204%20Feb%202014.pdf>. [Dátum prístupu: 15. marec 2025].
- [100] Gartner, "Kvalita údajov: osvedčené postupy pre presné informácie", 1. januára 2025. [Online]. Dostupné na: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "For Want of a Nail," [Online]. Dostupné na: [https://en.wikipedia.org/wiki/For\\_Want\\_of\\_a\\_Nail](https://en.wikipedia.org/wiki/For_Want_of_a_Nail). [Dátum prístupu: 15. marec 2025].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information", október 2013. [Online]. Dostupné na: [https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi\\_open\\_data\\_fullreport\\_oct2013.pdf](https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf). [Dátum prístupu: 15. marca 2025].
- [103] EY, "Cesta k uhlíkovej neutralite", 10. marca 2023. [Online]. Dostupné na: [https://www.ey.com/ru\\_kz/services/consulting/the-path-to-carbon-neutrality](https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality). [Dátum prístupu: 15. marec 2025].
- [104] PWC, "ESG Awareness", 1. júla 2024. [Online]. Dostupné na: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Dátum prístupu: 15. marca 2025].
- [105] G. Hammond, "Embodied Carbon - The Inventory of Carbon and Energy (ICE)", 2024. [Online]. Dostupné na: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO2\_calculating the embodied carbon", 2024. [Online]. Dostupné na: [https://github.com/datadrivenconstruction/CO2\\_calculating-the-embodied-carbon](https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon).
- [107] McKinsey, "Predstavujeme si digitálnu budúcnosť stavebníctva", 24. júna 2016. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Dátum prístupu: 25. februára 2025].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10. októbra 2024. [Online].

- Dostupné na: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Dátum oslovenia: 15. marec 2025].
- [109] SAS, "Dátové jazero a dátový sklad - poznajte rozdiel," [online]. Dostupné na: [https://www.sas.com/en\\_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html](https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html). [Dátum odkazu: 15. marec 2025].
- [110] ADSK, "Informačné modelovanie budov", 2002. [Online]. Dostupné na: [https://www.laiserin.com/features/bim/autodesk\\_bim.pdf](https://www.laiserin.com/features/bim/autodesk_bim.pdf). [Dátum oslovenia: 15. marec 2025].
- [111] A. Bojko, "Mapa história BIM", 2024. [Online]. Dostupné na: <https://bigdataconstruction.com/history-of-bim/>.
- [112] A. S. Borkowski, "Definície BIM podľa organizácií a noriem", 27. decembra 2023. [Online]. Dostupné na: <https://encyclopedia.pub/entry/53149>. [Dátum oslovenia: 5. marca 2025].
- [113] CAD Vendor, "OPEN BIM Programme", 2012. [Online]. Dostupné na: [https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open\\_bim/](https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/). [Dátum referencie: 30. marec 2025].
- [114] Wikipedia, "Industry Foundation Classes," [Online]. Dostupné na: [https://en.wikipedia.org/wiki/Industry\\_Foundation\\_Classes](https://en.wikipedia.org/wiki/Industry_Foundation_Classes). [Dátum prístupu: 15. marec 2025].
- [115] Wikipedia, "IGES," [Online]. Dostupné na: <https://en.wikipedia.org/wiki/IGES>. [Dátum odkazu: 30. marec 2025].
- [116] A. Bojko, "História CAD (BIM)", 15. decembra 2021. [Online]. Dostupné na: [https://miro.com/app/board/o9J\\_laML2cs=/](https://miro.com/app/board/o9J_laML2cs=/). [Dátum oslovenia: 24. februára 2025].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Referenčná štúdia softvérovej podpory IFC: GeoBIM benchmark 2019 - časť I", 8. 1. 2021. [Online]. Dostupné na: <https://arxiv.org/pdf/2007.10951.pdf>. [Dátum prístupu: 5. marca 2025].
- [118] I. Rogačev, "Hovorme o BIM: Maxim Nečiporenko | Renga | IFC | Domáci BIM," 13. apríla 2021. [Online]. Dostupné na: <https://www.youtube.com/watch?t=3000&v=VO3Y9uuzF9M&feature=youtu.be>. [Dátum vystúpenia: 5. marca 2025].
- [119] D. Ares, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth" (RETS v oblasti nehnuteľností: prečo je kľúčový pre efektívnosť a rast), 17. decembra 2024. [Online]. Dostupné na: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Dátum prístupu: 5. marca 2025].

- [120] "Flex token cost", 2024. [Online]. Dostupné na: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] A. Bojko, "Zabudnite na BIM a demokratizujte prístup k údajom (17. Kolokvium Investor - Hochschule - Bauindustrie)", 2024. [Online]. Dostupné na: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagment-und-Bauindustrie.pdf>.
- [122] Д. Hill, D. Foldesi, S. Ferrer, M. Friedman, E. Loh a F. Plaschke, "Solving the construction industry productivity puzzle", 2015. [Online]. Dostupné na: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industrys-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle," 1. januára 2018. [Online]. Dostupné na: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Dátum prístupu: 2. marca 2025].
- [124] Apple.com, "Pixar, Adobe, Apple a NVIDIA vytvorili Alianciu pre OpenUSD s cieľom presadzovať otvorené štandardy pre 3D obsah," 1. augusta 2023. [Online]. Dostupné na: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Dostupné: 2. marca 2025].
- [125] AECmag, "ADSK's granular data strategy", 25. júla 2024. [Online]. Dostupné na: <https://aecmag.com/technology/autodesks-granular-data-strategy/>. [Dátum prístupu: 15. marca 2025].
- [126] A. Bojko, "Vek zmien: IFC je minulosťou alebo prečo sa ADSK a ďalší dodávatelia CAD chcú vzdať IFC za USD v 14 kľúčových faktoch", 24. 11. 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Dátum oslovenia: 23. febr. 2025].
- [127] A. Bojko, "ENG BIM Cluster 2024 | Bitka o dátá a aplikácie LLM a ChatGPT v stavebníctve", 7. augusta 2024. [Online]. Dostupné na internete: ENG BIM Cluster 2024 | Bitka o dátá a aplikáciu LLM a ChatGPT v stavebníctve. [Dostupné: 15. marca 2025].
- [128] "Jeffrey Zeldman predstavuje", 6. mája 2008. [Online]. Dostupné na: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Dátum prístupu: 15. marec 2025].
- [129] A. Boyko, "DWG Analyse with ChatGPT | DataDrivenConstruction", 5. marca 2024. [Online]. Dostupné na: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Dátum odkazu: 15. marca 2025].

- [130] McKinsey , "McKinseyho sprievodca prekonávaním konkurencie vo veku digitálnych technológií a umelej inteligencie", 2023. [Online]. Dostupné na: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Dátum oslovenia: 30. marca 2025].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31. marca 2016. [Online]. Dostupné na: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Dátum prístupu: 15. marca 2025].
- [132] J. Bertin, "Grafika a grafické spracovanie informácií", 8. septembra 2011. [Online]. Dostupné na: [https://books.google.de/books/about/Graphics\\_and\\_Graphic\\_Information\\_Process.html?id=csqX\\_xnm4tcC&redir\\_esc=y](https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y). [Dátum prístupu: 15. marca 2025].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking," [online]. Dostupné na: <https://www.causeweb.org/cause/resources/library/r1266>. [Dátum prístupu: 15. marec 2025].
- [134] Ministrymagazine, "Ako veda objavila stvorenie", január 1986. [Online]. Dostupné na: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Dátum odkazu: 15. marec 2025].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23. mája 2017. [Online]. Dostupné na: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Dostupné 15. mája 2024].
- [136] "Ako vytvoriť dátovú architektúru, ktorá bude hnacím motorom inovácií dnes a zajtra", 3. júna 2020. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Dátum vystúpenia: 15. marca 2025].
- [137] Oxford, "Woodrow Wilson 1856-1924," [Online]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-ored6-00011630>. [Dátum vystúpenia: 15. marec 2025].
- [138] "Konvertori", 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1. januára 2017. [Online]. Dostupné na: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Dátum prístupu: 18. februára 2025].
- [140] "Potrubie vo výstavbe", 2024. [Online]. Dostupné na: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Wikipedia, "Apache NiFi", 1. januára 2025. [Online]. Dostupné na:

- [https://de.wikipedia.org/wiki/Apache\\_NiFi](https://de.wikipedia.org/wiki/Apache_NiFi). [Dátum prístupu: 5. marca 2025].
- [142] n8n, "Gmail AI Auto-Responder: Vytváranie návrhov odpovedí na prichádzajúce e-maily", 1. mája 2024. [Online]. Dostupné na: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Dostupné 15. marca 2025].
- [143] n8n, "Real Estate Daily Deals Automation with Zillow API, Google Sheets and Gmail", 1. marca 2025. [Online]. Dostupné na: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with-zillow-api-google-sheets-and-gmail/>. [Dátum prístupu: 15. marca 2025].
- [144] B. T. O'Neill, "Miera neúspešnosti projektov v oblasti analytiky, umelej inteligencie a veľkých dát = 85 % - fuj!", 1. januára 2025. [Online]. Dostupné na: <https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*, Oxford University Press, 1934.
- [146] T. J. S. a J. S. Jesse Perla, "A Problem that Stumped Milton Friedman", Quantitative Economics with Julia, 1. 1. 2025. [Online]. Dostupné na: [https://julia.quantecon.org/dynamic\\_programming/wald\\_friedman.html](https://julia.quantecon.org/dynamic_programming/wald_friedman.html). [Dátum prístupu: 1. mája 2024].
- [147] T. Landsall-Welfair, Predpovedanie aktuálnej nálady národa, Significance, 2012.
- [148] A. Bojko, "San Francisco. Stavebný sektor 1980-2019," 2024. [Online]. Dostupné na: <https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] A. Boyko, "Kaggle: RVT IFC Files 5000 Projects", 2024. [Online]. Dostupné na: <https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Príprava na budúcnosť s prepojenou výstavbou," [Online]. Dostupné na: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>. [Dostupné 15. marca 2025].
- [151] Cisco, "Prieskum Cisco odhalil, že takmer tri štvrtiny projektov internetu vecí zlyhávajú", 22. mája 2017. [Online]. Dostupné na: <https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html>.
- [152] "Podmienky potrebné na zachovanie fosílií rastlín", 2024. [Online]. Dostupné na: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "Fink z BlackRock o dlhopisoch, fúziách a akvizíciách, recesii v USA a voľbách: celý rozhovor,"

2023. [Online]. Dostupné na: <https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 slávnych katastrof umelej inteligencie", 02. októbra 2024. [Online]. Dostupné na: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>. [Dátum prístupu: 15. marca 2025].
- [155] G. Kasparov, Deep Thinking, PublicAffairs, 2017.
- [156] Wikipedia, "Kaggle", 1. januára 2025. [Online]. Dostupné na: <https://en.wikipedia.org/wiki/Kaggle>. [Dátum prístupu: 15. marec 2025].
- [157] Kaggle, "Titanic - strojové učenie z katastrofy", 1. januára 2025. [Online]. Dostupné na: <https://www.kaggle.com/competitions/titanic/overview>. [Dostupné 10. marca 2025].
- [158] Ш. Johri, "Vytváranie ChatGPT: od údajov k dialógu", 2024. [Online]. Dostupné na: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Domingos, "Niekol'ko užitočných vecí, ktoré treba vedieť o strojovom učení", 2024. [Online]. Dostupné na: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Saramago, "Citát," [online]. Dostupné na: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Dátum prístupu: 17. marec 2025].
- [161] NVIDIA, "Vylepšite svoje tréningové údaje pomocou nových modelov klasifikátora NVIDIA NeMo Curator", 19. decembra 2024. [Online]. Dostupné na: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Dostupné 25. marca 2025].
- [162] "NVIDIA oznamuje významné vydanie modelov a nástrojov fyzikálnych dát nadácie Cosmos World Foundation," 18. marca 2025. [Online]. Dostupné na: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Dátum prístupu: 25. marca 2025].
- [163] NVIDIA, "NVIDIA Isaac Sim," [na internete]. Dostupné na: <https://developer.nvidia.com/isaac/sim>. [Prístup 25. marca 2025].
- [164] M. Quarterly, "Prečo digitálne stratégie zlyhávajú", 25. januára 2018. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Dátum prístupu: 15. marca 2025].
- [165] M. J. Perry, "Moje oblúbené citáty Miltona Friedmana", 17. novembra 2006. [Online]. Dostupné na: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Dátum prístupu: 1. marec 2025].

- [166] J. A. Wheeler, "Informácie, fyzika, kvantá: hľadanie súvislostí", 1990.
- [169] A. Bojko, "Lobistické vojny a rozvoj BIM. Časť 5: BlackRock je pánom všetkých technológií. Ako korporácie ovládajú otvorený zdrojový kód," 2024. [Online]. Dostupné na: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [170] T. Krijnen a J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.
- [171] "Počet podnikov v sektore stavebníctva v Spojenom kráľovstve v roku 2021 podľa veľkosti podniku", 2024. [Online]. Dostupné na: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5 000 projektov IFC&RVT", 2024. [Online]. Dostupné na: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] M. Popova, "It from Bit: Pioneering Physicist John Archibald Wheeler on Information, the Nature of Reality, and Why We Live in a Participatory Universe", 2008. [Online]. Dostupné na: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Dátum odkazu: február 2025].
- [174] *Lobistické vojny o dátá v stavebníctve / Technofeudalizmus a história skrytej minulosti BIM*. [Film]. Nemecko: Artyom Boiko, 2023.
- [175] A. Boyko, "CHATGPT WITH REVIT AND IFC | Automatické vyhľadávanie dokumentov a údajov z projektov", 16. 11. 2023. [Online]. Dostupné na: [https://www.youtube.com/watch?v=ASXolti\\_YPs&t](https://www.youtube.com/watch?v=ASXolti_YPs&t). [Dátum prístupu: 2. marca 2025].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value", 22. 1. 2022. [Online]. Dostupné na: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Dátum prístupu: 15. februára 2025].
- [177] KPMG, "Stavebníctvo v digitálnom svete", 1. mája 2021. [Online]. Dostupné na: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Dátum prístupu: 5. apríla 2024].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17. marca 2023. [Online]. Dostupné na: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Dátum prístupu: 15. februára 2025].
- [179] O. Business, "Satya Nadella prezradil, ako agenti AI narušia modely SaaS", 10. januára 2025. [Online]. Dostupné na: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Dátum prístupu: 15. marca 2025].

2025].

- [180] Forbes, "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says," 23. marca 2016. [Online]. Dostupné na:  
<https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Dátum prístupu: 15. marca 2025].
- [181] Ministerstvo zahraničných vecí, Commonwealthu a rozvoja Spojeného kráľovstva, "Stratégia digitálneho rozvoja 2024 až 2030", 18. marca 2024. [Online]. Dostupné na:  
<https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030/digital-development-strategy-2024-to-2030>. [Dátum prístupu: 15. marca 2025].
- [182] "Vízia a stratégia v odvetví stavebného dizajnu", 7. novembra 2003. [Online]. Dostupné na:  
<https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Dátum vystúpenia: 5. marec 2025].
- [183] M. Bocharov, "Informačné modelovanie", marec 2025. [Online]. Dostupné na:  
<https://www.litres.ru/book/mihail-evgenevich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Dátum vystúpenia: 15. marca 2025].
- [184] "Integrated Design-Through-Manufacturing: Benefits and Rationale," 2000. [Online]. Dostupné na:  
[https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Dátum odkazu: 25. marec 2025].
- [185] CAD Vendor, "Program Open BIM je marketingová kampaň", 12. marca 2012. [Online]. Dostupné na: <https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Dátum oslovenia: 30. marca 2025].

# INDEX PREDMETOV

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287  
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

## A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Apache ORC, 62, 330, 378

Apache Parquet, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

## B

BDS, 258, 259

Veľké dáta, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Bokeh, 320, 337

BOM, 76, 77, 79, 257, 263

ohraničujúci rámček, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

## C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421

Claude, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

CO<sub>2</sub>, 72

CO<sub>2</sub>, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

RE, 56, 168, 169, 170, 477

Kopilot, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

**D**

DAE, 276, 277, 278, 280, 281, 284, 285, 414  
 DAG, 362, 363, 365, 366  
 Pomlčka, 320, 336, 337.  
 Správa údajov, 373, 395, 396, 398, 400, 401  
 Data Lake, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400  
 Data Lakehouse, 67, 373, 386, 387, 388  
 Minimalizmus údajov, 373, 395, 396, 400, 401  
 Data Swamp, 373, 395, 397, 401  
 Dátový sklad, 382, 383, 400  
 Údaje ako služba, 487  
 založené na údajoch, 50, 170, 460, 461, 484, 486  
 DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475  
 DataOps, 170, 373, 398, 399, 400, 401, 475  
 DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494  
 DGN, 8, 140, 186, 227, 357  
 DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497  
 DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390  
 DXF, 8, 73, 277

**E**

ECM, 58, 175  
 ECS, 142  
 EIR, 289  
 eLOD, 289  
 ELT, 384, 385  
 EPM, 14, 166, 198  
 ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492  
 ESG, 196, 235, 236, 238  
 ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494  
 Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497  
 Výpis, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

**F**

Perie, 62, 123, 330, 378  
 FPDF, 339, 340, 341, 342, 343

**G**

GDPR, 109  
 GIS, 58  
 GLTF, 143, 278  
 Tabuľky Google, 368, 370  
 Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

**H**

HDF5, 62, 67, 123, 329, 330, 378, 379, 380  
 HiPPO, 29, 37, 95, 424, 477, 484, 490  
 HTML, 123, 340, 365, 370

**I**

IDS, 289, 290, 291.  
 IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497  
 IGES, 262, 263, 276.  
 iLOD, 289  
 IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484  
 ISO 19650, 388

**J**

JavaScript, 320, 369, 378  
 JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480  
 Zápisník Jupyter, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

**K**

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433  
 k-NN, 392, 393, 442, 445, 446, 447, 448  
 KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

**L**

LEED, 235, 236, 238.  
 LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494  
 LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103,  
 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114,  
 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134,  
 135, 166, 187, 219, 220, 221, 222, 222, 224,  
 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303,  
 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334,  
 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348,  
 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413,  
 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461,  
 476, 488, 494  
 Náklad, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334,  
 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384,  
 450, 475  
 LOD, 287, 289  
 LOI, 287  
 LOMD, 287  
 Low-Code, 368, 369

## M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411,  
 494  
 MCAD, 77, 78, 257, 284  
 MEP, 14, 175, 492  
 SIETE, 234, 283, 284, 285, 296, 462  
 Microsoft SQL, 65  
 Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300,  
 329, 343, 349, 381, 421, 429, 443, 494  
 MRP, 11, 12  
 MS Project, 70  
 MySQL, 63, 64, 65, 291, 329, 330

## N

n8n, 116, 311, 361, 368, 369, 370, 371, 372  
 NLP, 69, 394  
 No-Code, 368, 369  
 NURBS, 142, 282, 283, 284, 285  
 NWC, 8, 276

## O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285  
 OCCT, 273  
 OCR, 69, 128, 132, 134  
 Trieda OmniClass, 154, 155, 156.  
 otvorený BIM, 142, 216, 219, 256, 261, 278, 291  
 Otvorený zdroj, 45, 55, 97, 98, 107, 108, 273, 275, 494  
 SOVA, 267, 268, 269.

## P

Pandy, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125,  
 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300,  
 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409,  
 410, 411, 412, 414, 475, 479, 494  
 Parkety, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382,  
 386, 401, 403, 414, 474, 480  
 PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132,  
 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338,  
 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355,  
 356, 357, 365, 376, 475, 480, 497  
 PDM, 16  
 PHP, 63, 339  
 Potrubie, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312,  
 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430,  
 476  
 PLM, 16, 246  
 PLN, 8, 272, 296  
 Plotly, 320, 336, 337, 372  
 PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246,  
 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390,  
 458, 462, 487  
 PMS, 84, 151, 233  
 PostgreSQL, 63, 64, 65, 329, 395  
 Power BI, 320, 372, 482  
 súkromný kapitál, 464, 487  
 Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129,  
 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224,  
 225, 227, 303, 304, 320, 329, 330, 332, 333, 339,  
 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426,  
 459, 475, 476, 479, 494

## Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225,  
 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479  
 QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329,  
 349, 343, 349, 381, 429, 443

## R

RAG, 111, 116  
 RDBMS, 63, 64, 65, 82, 89.  
 RDF, 267, 268, 269  
 RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474  
 RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482  
 ROI, 311, 317, 319, 321, 351, 370, 389, 479  
 RPM, 14, 84, 151, 331, 333, 334  
 RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329,  
 357, 376, 414, 415, 417, 497

## S

SaaS, 24, 50, 51, 52, 458  
ROZSAH, 273, 277  
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414  
Seaborn, 123, 320, 336, 337, 408, 412  
SPARQL, 269  
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494  
SQLite, 63, 64, 65, 166, 167, 296, 329, 363  
KROK, 261, 262, 263, 266, 272, 276, 277, 292  
Streamlit, 336, 337  
SVF, 142, 276, 285

## T

Transformácia, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

## U

Uniclass, 154, 155, 156.  
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

## V

VectorOps, 373, 398, 400, 401  
VR, 84, 271, 285

## W

WhiteBox, 240, 242, 243.

## X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480  
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474